THE ROLE OF SPANINS IN OUTER MEMBRANE DISRUPTION

A Dissertation

by

RAMCHANDER ROHIT KONGARI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ryland Young
Committee Members,	James Hu
	Tatyana Igumenova
	Steve Lockless
Head of Department,	Gregory Reinhart

August 2018

Major Subject: Biochemistry

ABSTRACT

Spanins are bacteriophage lysis proteins responsible for disruption of outer membrane (OM), the last step of phage lysis in Gram-negative hosts. The absence of spanins results in a terminal phenotype of fragile spherical cells. Spanins can be broadly classified into two types: two-component spanins (e.g., λ Rz-Rz1) and unimolecular spanins (e.g., T1 gp*11*). The former consists of two proteins, an integral inner membrane (IM) protein and OM lipoprotein, designated as i-spanin and o-spanin respectively. The two-component spanins interact through their C-termini to form a spanin complex spanning the periplasm and connecting the membranes. In contrast, unimolecular spanins (u-spanins) have an N-terminal OM lipoprotein signal and a C-terminal transmembrane domain (TMD). After maturation and localization, the u-spanin spans the entire periplasm as a single molecule, connecting the membranes with a covalent polypeptide chain. Our model for function of both types of spanins is that destruction of the peptidoglycan (PG) by the endolysin liberates the spanins to diffuse laterally and oligomerize, providing free energy for conformational changes that result in fusion of the IM and OM.

. The aim of the research presented here was to a) expand our understanding about two-component and u-spanins through a comprehensive bioinformatics analysis and b) investigate the molecular mechanism of u-spanin function. The bioinformatics study provided insights into spanin function, evolution, and domain organization. In the molecular analysis of gp*11*, the necessity of membrane localization signals of for function was validated using site-directed mutagenesis. Studies using a C-terminal GFP tagged allele of gp*11* indicated localized oligomerization within the PG meshwork. Several regulatory elements, including a truncated "anti-spanin" gene product, were identified. Furthermore, single missense mutations defective in

lysis found to be distributed throughout the periplasmic domain of gp*11* and shown to exhibit oligomerization defects. The results from a spheroplast fusion assay designed to test our model of gp*11* function are also discussed. In summary, this work demonstrates that spanins, like viral membrane fusion proteins, adopt different strategies to achieve fusion of the inner and outer membranes.

DEDICATION

To my late grandmothers Balamma Kongari and Kanthamma Chinthakayala, who passed away during my journey of becoming a doctorate. It still pains me that I could not be with them when they breathed their last. While they might be gone, their love and blessings will always continue to be with me.

ACKNOWLEDGEMENTS

I would like to start by thanking my committee chair Dr. Ryland Young for giving me this opportunity to work under his guidance. He is easily the best mentor that I have ever had, both in professional and personal terms. There was never a single meeting with Ry that I did not end up being more motivated than before, more satisfied about my work than I was before. I definitely owe him a lot for not only giving me the freedom to do whatever I did, but also supporting me through each and every phase of my Ph.D. He showed confidence in me at times when I was almost close to giving up. All the awards and recognitions I received would not have been possible without his encouragement and endorsement. He's been such a strong influence that, I hope to one day replicate his teaching, leadership and management styles and may be, be as excited about science and have a sense of humor as him at that "young" age. Still, how much ever I try, I will never be able to beat his attention to detail and his eye for grammatical/spelling errors.

Next, I would like to thank my committee members, Dr. James Hu, Dr. Tatyana Igumenova, and Dr. Steve Lockless for continuously encouraging me, providing insightful inputs for my research and guiding me towards excellence. Dr. Hu, who was also my ORP mentor in the first year, is someone who I have had some of the best science discussions with. I thank Dr. Igumenova, with whom I had the privilege of taking 2 courses and a rotation in my first year, for keeping me focused and advising on prioritization. I thank Dr. Lockless for always pushing me to think ahead of time and about different perspectives of research, and all the valuable advice on my career path.

Coming to the Young Lab, I would like to first thank Dr. Catrina Reed, who was my mentor during my rotation and introduced me to the "Young Lab" way of research. I really

appreciate Dr. Joel Berry, Dr. Manoj Rajaure (the Guru) and Dr. Jesse Cahill, the spanin core group, for giving me continuous inputs on my research, new ideas for experiments, and critique on all the spanin manuscripts. Thanks to Karthik for making me more organized and also a morning person, which I am sure will be very helpful in the long run. Adriana has been more like the sister from another mother that I found here, she's always been there to lend an ear whenever I had to pour out any frustration, in both professional and personal aspects, always being supportive and bringing me back to reality. This Ph.D. would not have been possible without the clerical assistance of Ms. Daisy Wilbert, my go to person for literally any need and the most important person behind all Young Lab operations. I have had the luck of mentoring quite a few talented undergrads and rotation students, all of who have made significant contributions to my work, particularly the efforts from Jacob Simpson, Anthony Pratt, Jeffrey Snowden, Jessica Tracy and Casey Tatsch yielded great results. I would also like to thank all the other undergrads in the lab, especially Chandler for dealing with all those last minute requests for media and other stuff, and also a special shout out to Lorna for dealing with my sense of humor.

I would like to thank all the members of CPT, the grad students and faculty alike. I always got valuable feedback at the lab meeting presentations and learned so much from others as well. I am especially proud of our Journal Club participants, for the way we took over control and made it even better than what it was before. I would also like to sincerely thank Dr. Qiuyan Shao for all the help with the microscopy. My classmates Claudia, Jeremy, Karl, Keya, Ting, Xiaoyuan, Yi and Yuan have been great friends and I will cherish all our moments together. The BGA has really been a family away from home. Having been a BGA officer officially for four years (unofficially all the time), I am really proud of the way the grad students in our department take the initiative, try to improve the quality of life, and have succeeded in implementing so

many changes in departmental policies for the everyone's good. Dan, Kristina, Alyssa, Reed, Alfredo, Julian, Andrew, Jeremy (Weaver), Daniel (Shoup), Joel Gray, Sarah, Mehmet, Callie, Dakota, and so many others that I am currently too emotional to remember.

Cheers to the folks from ISMA, GPSC, SRW, Grad camp, BUILD, and FLC. I would also like to thank the members of the Indian Graduate Student Association, which I have been a part of for all my stay at Aggieland, for the life-long friendships and unforgettable memories. I would particularly like to thank the team that I led as the President, all of who have continued to be my best friends for all the crazy and fun times. It was their nominations for the Aggie Core Values award that got me the Aggie ring which I proudly wear every day (and surely could not have afforded!).

I would like to thank Anuj and Biren who have been great friends, amazing mentors, unspoken brothers and helped me out in a lot of different ways. Anuj and I co-authored a paper on a research idea that was literally discussed over "buffer B" at 2 AM. I will greatly miss the HQ times with Apoorv, Adi, Supriya, Bhavya, Lal, Swoop, Sandy, Rohan, K, G, Lel and Hemi! Thanks to Akshaya for all the man hours she spent on me, I am not sure how she manages to be my strongest supporter and harshest critic at the same time. I have to thank Noor for introducing me to the field of architecture, teaching me so many things about introverts, for the "hopelessly-hopeful tag", the random fun-filled road trips, the deep discussions, and so many other things. Also, Navya for dealing with all the puns and bad humor, and Nadiya for being my snap and fitbit buddy!

I would like to thank all my roommates from 101C, Maple Ave for the amazing discussions, great food and their friendship. Don, Loki, AK, DJ, Sameer, Soma, AJ, Neeraj (baba), Waseem and Shravan (my personal hype men!), thanks for being part of this journey. I

would also like to thank the Gult boys, especially Yati, Narri, Vishnu, Chitti, TJ and Swaro, thanks for our discussions about music. basketball, and the times we played racquetball. I would like to acknowledge the Cherry folks Vasav, JD, Shweta, and Shreya for all the great food, fun and maalbro times.

The Hydies, my friends from undergraduate time at NITW, for everything. Every time I hit a low, these guys were always there to pick me up and reminded me I always had someone to turn to, whatever might be the situation. What started off as a joke, the phrase "Mama, sab khairiyat?" was so important for me, in the toughest of times. It would not be an exaggeration to state that they wanted me to get this doctorate degree more than me, and this would not have been possible without their constant push, encouragement, and support! Thanks for sticking with me at all times guys, Friend to friend, brother to brother!

I would not like to thank the US Homeland Security department for putting me through the 221g administrative processing every time I attended a visa interview. If not for them, I might have been done a year earlier.

Finally, I would like to thank my family, my dad for giving me the freedom to pursue my dreams abroad, and my mom, who has painfully stayed apart from me for the past 8 years. Without any major educational background, my parents encouraged and supported me all the time, and showed confidence in me irrespective of the situations. I would also like to thank my sister, Maddy from the bottom of my heart for holding down the fort back home while I was gone. You are one of the strongest women that I have known, and I love you for that! I need to thank the newest addition to our family, my brother in law, Sai Krishna for effortlessly becoming a close friend and being so understanding in all situations. I would also like to thank my extended family, especially all the kids in the house for showering much love on me all the time.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

The research and writing of this work were primarily conceived and carried out by Rohit Kongari (RK) under the supervision of a dissertation committee consisting of Dr. Ryland Young (advisor), Dr. James Hu, and Dr. Tatyana Igumenova of the Department of Biochemistry and Biophysics, and Dr. Steve Lockless of Department of Biology. Dr. Ryland Young. Dr. Manoj Rajaure (MR) and Dr. Joel Berry (JB) contributed to creating and populating the SpaninDB. MR performed the experiments for the cysteine mutational analysis of T4 spanins. Eric Rasche and Eleni Mijalis were responsible for development and maintenance of various tools on the CPT Galaxy instance used for analysis in Chapter II, as well as the online version of the SpaninDB. Dr. Jesse Cahill, along with MR and JB, contributed to the interpretation of results and provided critical feedback on the manuscript. Multiple undergraduate researchers and graduate rotation students also contributed to the experiments performed in Chapter III and IV.

Funding sources

This work was supported by Public Health Service grant GM27099, and by the Center for Phage Technology at Texas A&M University, jointly sponsored by Texas AgriLife.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Bacteriophages: history and diversity

Bacteriophages, also simply referred to as phages, are viruses that infect and kill bacteria. Phages are essentially composed of proteins that encapsulate a nucleic acid genome. At a global population of ~10^{31}, they are the most abundant organisms on the biosphere [1, 2]. Phages can be found in all places where their host bacteria are present, in environments ranging from soil to water [3-5], mammal guts to sewage plants[6, 7], deserts to ocean beds[8, 9], and hot springs to arctic cores[10, 11]. While preying on and coexisting with their hosts, phages have a huge impact on the bacterial population dynamics around them [12-14]. For example, phages account for the turnover of about 40% of the ocean bacteria each day, and thus, play a major role in the global carbon cycle [2, 15]. In addition, phages also contribute to the bacterial diversity and evolution through horizontal gene transfer, carrying toxin genes and other virulence factors [16-18]. Phages are extremely diverse; a few phages have genomes that encode as few as four genes enveloped inside a simple icosahedral protein structure made up of a single coat protein, but there also are phages coding for nearly a thousand genes and with oblate and highly decorated capsids [19]. Phage genomes are also highly mosaic; i.e. each genome represents a different and unique combination of the same modules that have been shuffled and re-shuffled over billions of years of viral evolution [20, 21]. Phage genomes likely represent the largest reservoir of unexplored genes and proteins anywhere in the biosphere. Modern molecular biology and biotechnology have been greatly impacted by bacteriophages through the contributions from phage research. A lot of scientific breakthroughs like identification of nucleic acids as the

hereditary component, the discovery of genetic recombination, understanding how the "genetic-code" is read, the fundamental phenomena behind DNA replication, transcription and translation, the discovery of restriction enzymes etc. to name a few, were all outcomes of experiments involving phages.

The bactericidal activity of phages was reported for the very first time by Ernest Hanbury Hankin in 1896 [22]. He noticed that there was "something" in the water of the holy rivers Ganges and Yamuna in India, that could pass through very fine porcelain filters and kill *Vibrio cholerae*. Bacteriophages were discovered in 1915 by an English Microbiologist Frederick Twort (1877-1950) [23], and two years later, "independently" by a French-Canadian self-trained microbiologist Felix d'Hérelle (1873-1949) [24]. Controversies still exist about whether d'Herelle was completely honest or not about his knowledge of Twort's discovery in 1915 when he published his own work in 1917 [25]. Twort believed the agent responsible for bacterial killing activity could be an enzyme secreted by the bacteria themselves. However, the start of world war I and the loss of a grant halted any further investigations by Twort [26]. D'Herelle, while working at the Institute of Pasteur in Paris, noticed a cleared zone on one of his bacterial agar cultures. Upon mixing a filtrate of this cleared zone with a fresh sample of dysentery culture, he observed that the dysentery bacteria were quickly destroyed [24]. Unlike Twort, d'Herelle had little doubt about the nature of the phenomenon and proposed that it was caused by a virus capable of parasitizing bacteria, coining the term "bacteriophage".

Phage particles were purified for the first time by a Hungarian biochemist, Max Schlesinger, in the mid-1930s [27]. His chemical analysis revealed that phages were 50% protein and 50% DNA by mass. A few years later, electron micrographs of bacteriophages were produced by T. F. Anderson and Salvador Luria [28], who showed phage preparations exhibited

uniform morphology and size. However, extensive studies on phages and their role in laying the foundation of modern molecular biology were pioneered by a German physicist, Max Delbrück, working at Caltech [29]. Using the reductionist approach, Delbrück focused on phages, which he believed were the simplest biological entities. With the help of his colleagues, Al Hershey and Salvador Luria, Delbrück established an annual summer phage course taught at the Cold Spring Harbor Laboratory on Long Island [30]. Through this course, they together influenced a generation of researchers to be followers of "The Phage Church" and made everyone agree to focus on a pre-decided set of *E. coli* phages, designated as T1 through T7. This started what is considered to be the "Golden Era of Phage Biology" in the early 1940s, which lasted till the early 1980s. Delbrück, Hershey, and Luria were recognized with the Nobel Prize in 1960 for their contributions.

Phage therapy, the use of phages as therapeutic agents, was widely in practice between the 1920s and 1950s, before the advent of the antibiotics [26]. It was d'Hérelle who introduced the concept of phage therapy in 1919 when he successfully restored good health to a 12-year old boy with severe dysentery through the administration of phage preparations. d'Hérelle was fairly rewarded for his discovery of phages and was appointed as a Professor at Yale University Medical Center while continuing as a staff member at the Pasteur Institute. Starting with a commercial laboratory in Paris that produced several phage preparations against bacterial infections, he went on to establish phage therapy centers in several countries, including the U. S., France, and Soviet Georgia. Phage therapy was even used on wounded soldiers in World War II, more predominantly by the Soviet Union. Commercial production of phage lysates for therapy in the United States, Europe, and Russia continued up until the advent of antibiotics. The discovery of antibiotics and the subsequent intensive use led to a waning interest in phage therapy around

the world, especially the western hemisphere. Additionally, the sociopolitical image of phage therapy for being associated with researchers from communist background and lack of proper experimental protocols and documentation also contributed to the decline of phage therapy globally. Barring the Bacteriophage Institute in Tbilisi (Republic of Georgia), almost all the research organizations that showed an initial interest in the potential of phage therapy abandoned any further explorations.

However, due to the rise in the number of multiple drug resistant (MDR) bacterial species over the past few decades and the dwindling number of new antibiotics, phage therapy seems to be making a comeback. Phage cocktails have been shown to be effective against antibiotic-resistant *P. aeruginosa* infections of the skin, lungs, and gastrointestinal tract in mouse models [31]. Recent animal studies also demonstrated promising results for MDR *E. coli* [32], *V. parahaemolyticus* [33], *S. aureus* [34] and *A. baumanii* [35]. The high specificity of bacteria-phage interactions and the ease of generating phages that can overcome bacterial resistance are major driving factors in advocating for phages as a substitute and/or supplement to conventional antibiotics. With potential applications in not just human medicine, but also veterinary sciences, agriculture and other industries, phage therapy is gaining genuine consideration in different fields.

Phages can be classified in multiple ways. Based on the way phages carry their genetic information, they can be categorized into ribonucleic acid (RNA) and dideoxy ribonucleic acid (DNA) phages, which can be further divided into single-stranded (ss) and double-stranded (ds). Our knowledge about RNA phage biology is comparatively less established than that of the DNA counterparts, partly due to the limited availability of RNA phages to study. Only 12 ssRNA and 5 dsRNA phages have their complete genomes deposited in the NCBI Viral Genomes database

till date. According to the ninth International Committee on Taxonomy of Viruses (ICTV) report in 2011 [36], only two RNA phage families had been officially recognized: the *Leviviridae* (ssRNA phages), and the *Cystoviridae* (dsRNA phages). Attempts to increase the existing RNA phage diversity using novel phage hunting strategies and metagenomic data mining techniques are underway [37]. On the other hand, complete genomic sequence information is available for more than 80 ssDNA phages and 2000 dsDNA phages in GenBank. The majority of the dsDNA phages fall into the order of Caudovirales, the class of tailed-phages [38]. Bearing a single dsDNA chromosome, these phages are equipped with a tail that attaches to the host exterior and is engineered to work at almost a 100% efficiency of injecting the phage DNA into the host cytoplasm [39]. Once injected into the host, the phage DNA either stays linear or circularizes before further replication and can even transition between linear and circular forms. The Caudovirales are classified based on their tails into sipho, myo and podo phages [38]. Sipho phages have a long, flexible and non-contractile tail, while myo phages have a long non-flexible, contractile tail. Podo phages, on the other hand, consist of a short, non-contractile tail. Phages are also classified into virulent or temperate phages, based on their mode of replication [14]. Upon successful infection and entry of the viral DNA into the host cytoplasm, a cascade of phage-encoded molecular processes occurs that can culminate in either lysis or lysogeny. Virulent phages follow the lytic life cycle strictly, whereas temperate phages contain a genetic switch that allows them to choose between the lytic and lysogenic pathways based on the infection conditions [40, 41]. Lysogeny allows for the replication of the phage genome along with the host as a prophage [42]. In this state, the phage is harmless to the host until it is induced by an environmental stress signal to resume the lytic pathway. Induction also occurs spontaneously at a low frequency by unknown mechanisms. [43].

The bacterial cell envelope

To ensure progeny release, all phages must subvert the host cell envelope. The cellular envelope is a complex multilayered structure that serves multiple purposes [44]. It not only protects bacteria from their unpredictable and often hostile environment but allows for the exchange of selective cargo like nutrients, salts etc. The cell envelopes of most bacteria can be classified into two groups: Gram-negative or Gram-positive, based on a staining procedure developed by Christian Gram [45]. Gram-negative bacteria are surrounded by a thin peptidoglycan cell wall, which itself is surrounded by an asymmetric lipid bilayer known as the outer membrane. Gram-positive bacteria lack the outer membrane but possess a peptidoglycan layer many times thicker than that of the Gram-negatives. In addition, the Gram-positive cell envelope consists of long anionic polymers called teichoic acids, which thread out through the peptidoglycan. With regards to the experimental work described in this dissertation, the features of the cellular envelope of the Gram-negative host *E. coli* will be discussed in extensive detail.

Structure and biosynthesis of Gram-negative cell envelope

For phages that infect Gram-negative bacteria, the cellular envelope acts as a three-level physical barrier preventing progeny release at the end of the infection cycle. The structural components that constitute the three levels are the cytoplasmic or inner membrane (IM), the peptidoglycan (PG), and the outer membrane (OM) (Fig. 1.1).

The first structural obstacle to lysis is the IM. The IM is a semi-permeable, symmetric phospholipid bilayer composed of phosphatidyl ethanolamine, phosphatidyl glycerol, lysophosphatidyl ethanolamine, small fractions of cardiolipin and phosphatidyl serine [46, 47]. The IM encompasses the cytoplasm and is the site of energy production, protein secretion and

transport and lipid biosynthesis [44, 48]. The cell maintains a 0.2 V electrochemical gradient between the IM and the periplasm that provides the major source of cellular energy during aerobic metabolism. If the membrane potential across the IM is nullified, the cell becomes metabolically inactive and there is no net macromolecular synthesis.

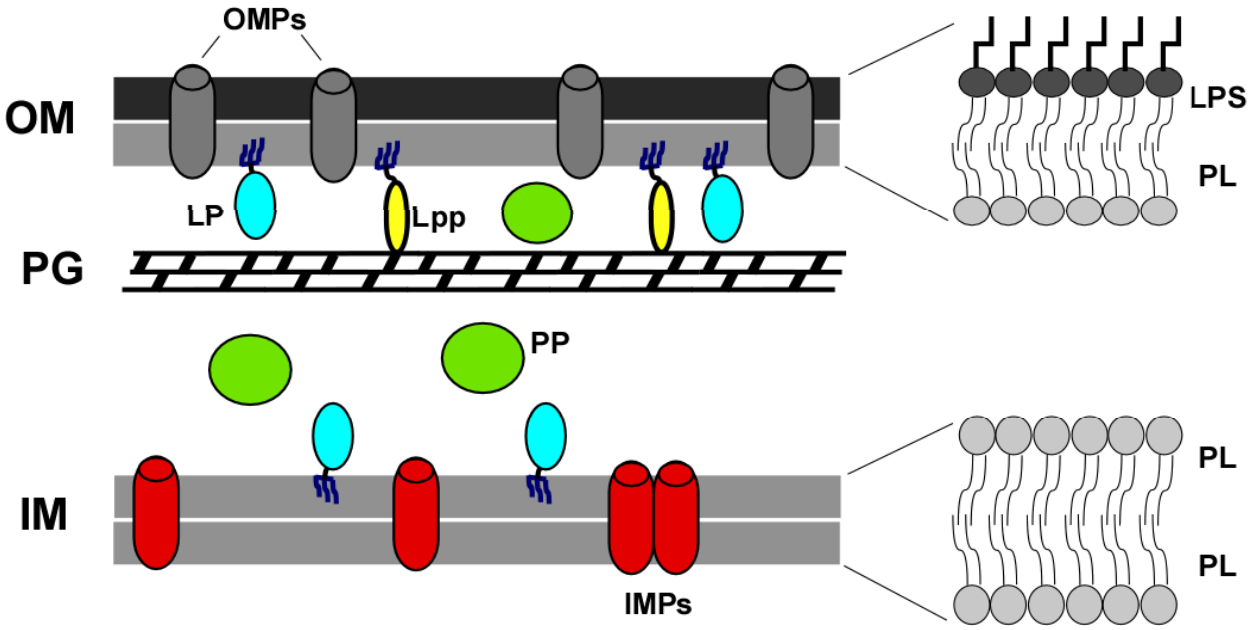


Figure 1.1 The *E. coli* cellular envelope. See text for details. Different components of the cellular envelope are shown and labeled using abbreviations as follows: IM, inner membrane; PG, peptidoglycan; OM, outer membrane; LPS, lipopolysaccharide; PL, phospholipid; OMPs, outer membrane proteins (Grey); IMPs, inner membrane proteins (Red); PP, periplasmic protein (Green); LP: lipoprotein LP (Cyan); Lpp, Braun's lipoprotein (Yellow). Insets show IM as a symmetric bilipid layer containing PL, whereas OM is an asymmetric bilayer with LPS on the outer leaflet and PL on the inner leaflet. Lpp is shown connected to the PG through a covalent linkage.

The periplasm constitutes the space between the IM and OM. The width of the periplasm depends on growth conditions and measurements using different sample preparation procedures have been documented [49-51]. Cryo-EM measurements on *E. coli* cells grown as per standard growth conditions used by our lab suggest the periplasmic width to be 25.5 ± 3.1 nm [52]. It has

also been demonstrated that the periplasm maintains the same osmolality as that of the cytoplasm [53]. The PG, which is composed of covalently linked glycan strands, is located in the periplasmic space. The glycan strands consist of repeating units of the disaccharide N-acetylglucosamine (NAG): N-acetylmuramic acid (NAM), linked together by β-1,4 glycosidic bonds [54]. On an average, these glycan strands consist of about 21 disaccharide units [55]. The glycan strands are further cross-linked by short peptides attached to the NAM units, forming a closed net-like structure surrounding the IM, also referred to as the murein sacculus [54]. The PG determines the shape of the cell and confers mechanical strength to resist internal turgor pressure [56]. The periplasm also contains many soluble proteins, including nucleases, chaperones, transport proteins, and proteases [57]. Another important component of the cellular envelope found in the periplasmic space are the membrane-derived oligosaccharides (MDO), which are generated to maintain the osmotic balance between the cytoplasm and the periplasm [58]. The MDOs are usually made up of 8-10 highly branched glucose units and are above the permeability limit of the OM [59].

In contrast to the IM, the OM is an asymmetric barrier, with an inner leaflet of phospholipids and an outer leaflet of lipopolysaccharides (LPS) [60]. Serving as the first line of defense against external agents, the presence of OM makes the Gram-negative bacteria to less susceptible to detergents and drugs compared to their Gram-positive counterparts. The stability of the OM is, to an extent, dependent on the negatively charged LPS and can be enhanced by the presence of millimolar concentrations of divalent cations [61]. Divalent cations compensate for the anionic character of both phosphate and carboxyl groups in the LPS and possibly participate in the formation of ionic bridges between adjacent LPS molecules [62]. Indeed, it has been shown that treatment with a cation-chelating agent such as ethylenediaminetetraacetic acid

(EDTA) led to increased sensitivity to hydrophobic antibiotics and lysozyme in *E. coli* and other Gram-negative bacteria [61]. The PG is covalently linked to the OM by an OM lipoprotein, Lpp, via a C-terminal Lys residue [63]. About one-third of the million copies of Lpp per cell is bound to the PG and the covalent linkage is important for the strength of the OM [44]. In addition, the OM consists of proteins known as porins that allow passive diffusion of molecules <700 Da [64]. These act as a means of providing a selective permeability barrier and limit the passage of periplasmic osmolytes involved in maintaining the osmotic balance across the OM.

Membrane protein processing and sorting

There are several different types of proteins that localize to different segments within the cellular envelope. Both the inner and outer membranes have integral membrane proteins with distinct secondary structures. All IM proteins span the lipid bilayer by alpha-helical transmembrane domains (TMDs) [65] whereas the integral membrane proteins of the OM span the membrane with beta-barrel structures [66]. All integral membrane proteins are synthesized in the cytoplasm and must be translocated or flipped across the IM to reach their final destination. Proteins destined for the periplasm or the OM are made initially in precursor form with a signal sequence at the amino terminus that targets them for translocation from the cytoplasm [67]. This translocation reaction is carried out by an essential, heterotrimeric IM protein complex called SecYEG [68]. The need for a signal sequence and a SecYEG like membrane protein complex for translocation across membranes is a conserved feature in both prokaryotic and eukaryotic systems [69]. The translocation reaction is driven by interactions of the SecYEG complex with the essential ATPase SecA [70]. Periplasmic and OM proteins are generally translocated in post-translational fashion; i.e., synthesis and translocation are not coupled [44]. It is believed that

SecYEG cannot handle folded molecules; thus, proteins must be secreted in a linear fashion from the amino to the carboxy terminus like “threading a needle”. Proteins are maintained in their unfolded form by the cytoplasmic SecB chaperone until they can be secreted [71]. After secretion across the IM, the signal sequence is proteolytically removed by Signal Peptidase I [72]. Other components of the Sec translocon, such as SecD, SecF, and YajC, perform important but nonessential function(s) during translocation, perhaps facilitating the release of secreted proteins into the periplasm. Various chaperones function to prevent misfolding and aggregation of the proteins newly released into the periplasm. For example, almost all the periplasmic proteins with a requirement of a disulfide bond for folding and/or function require members of the host *dsb* system, like DsbA, DsbB, DsbC and DsbG [73]. DsbA and DsbB are involved in the oxidation step to form disulfide bonds, whereas DsbC and DsbG are responsible for isomerization or reduction of non-native disulfides. Translocation of folded proteins requires the TAT (Twin Arginine Translocation) system [74]. The signal sequences recognized by the TAT system are similar to those of the Sec system but are longer and have a signature TAT motif: Z-R-R-x-*-*, where Z and “*” represent polar and hydrophobic residues respectively [75]. However, transport by the TAT system is a low throughput and energetically expensive process, thus availed by only a few proteins.

Another distinct class of membrane proteins found in both the membranes is the lipoproteins. Lipoproteins undergo a different mode of processing and are dependent on the Lol transport machinery (Fig. 1.2) [76, 77]. Lipoproteins are also secreted across the IM through the Sec translocon, as precursors carrying an N-terminal signal sequence ending in a lipobox. The last residue of the lipobox, a Cys residue, then gets modified with a diacyl glycerol moiety by Lgt (lipoprotein diacylglyceryl transferase). But instead of processing by SPaseI, lipoylation

signal sequences are cleaved at the lipobox Cys residue by Signal Peptidase II (SPaseII). An additional acyl group is added to the newly exposed N-terminal Cys residue by Lnt (apolipopotein transacylase) as the final step in lipoprotein maturation.

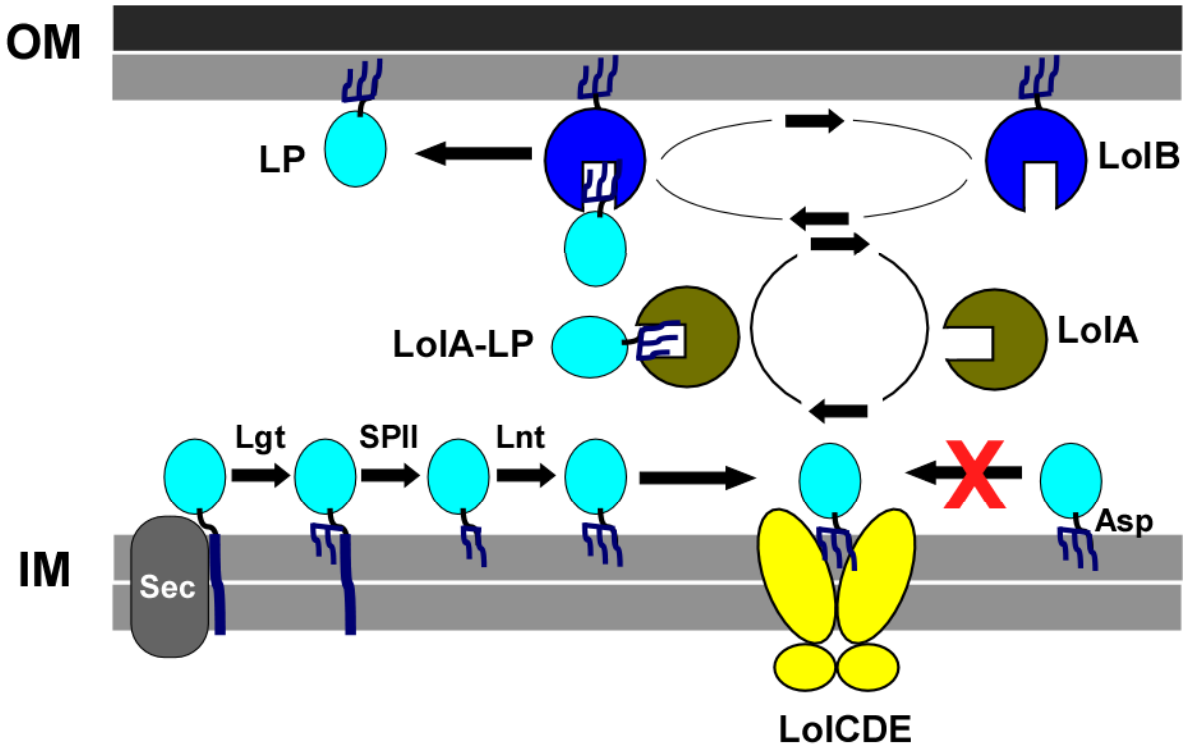


Figure 1.2 Lipoprotein processing and translocation. The different steps of lipoprotein processing and transport are shown, and the involved players are labeled. The lipoprotein, LP (Cyan) is translocated across the IM by the Sec translocon (grey). A diacyl-glycerol is added to the lipoylation cysteine by Lipoprotein glyceryl transferase (Lgt) and the signal sequence (dark blue) is cleaved by the Signal Peptidase (SPII). The N-terminal Cys of the processed lipoprotein is modified with one more acyl group by Apolipoprotein transacylase (Lnt) to form the mature lipoprotein. Lipoproteins destined for the OM are released from the IM by an ATP-binding cassette transporter, LolCDE (yellow). The released lipoprotein interacts with a soluble periplasmic chaperone, LolA (green) and is translocated to the OM as a complex (LolA-LP). The complex then delivers the lipoprotein to an OM receptor LolB (violet), which eventually transfers it to the periplasmic leaflet of the OM. Asp residues at the +2 and +3 positions of the mature lipoprotein, avoid the LolCDE ejection step.

Mature lipoproteins destined for the periplasmic leaflet of the OM are first released from the IM by an ATP-binding cassette transporter, LolCDE [78]. The release from the IM requires ATP hydrolysis and is immediately followed by the formation of a water-soluble complex between the released lipoprotein and the periplasmic chaperone LolA [79]. The LolA-lipoprotein complex navigates across the periplasm to reach the OM. The complex then delivers the lipoprotein to the OM receptor LolB, which itself is a lipoprotein [80, 81]. This interaction eventually leads to lipoprotein transfer to the periplasmic leaflet of the OM [76]. IM lipoproteins carry Asp residues at the +2 position (and sometimes also at the +3 position) of the mature lipoprotein and avoid the LolCDE ejection step [82].

Overview of the phage infection cycle

The phage infection cycle starts with the adsorption of the phage to the host cell. In *Caudovirales*, the adsorption machinery dedicated to specific host recognition is localized at the tail-end. Different adsorption strategies engage different components ranging from a simple tail tip to a complex base plate, and sometimes even involve the tail spikes and tail fibers [83]. The process of a bacteriophage attaching to its host cell is a combination of physical diffusion of the phage particle across the cell surface, interactions between the phage adsorption apparatus and cell surface receptors, and interaction-induced conformational changes in binding domains [84]. Tail proteins of phages infecting Gram-negative hosts are capable of recognizing a highly diverse set of host surface components, including surface proteins, polysaccharides and lipopolysaccharides [85]. The second stage of the infection cycle is the penetration step. This is mediated by the tail sheath, which acts like a hypodermic needle to inject the viral genome

through the cellular envelope [86]. The empty phage head and other components remain outside the bacteria while the genomic material is delivered into the cytoplasm.

Once the phage genome enters the host cell, gene expression is initiated by the host transcription machinery. In the case of the paradigm phage λ, the genes can be divided into three classes based on the stage of their expression: immediate early genes, delayed early genes, and late genes [87, 88]. The immediate early genes are under the control of promoters recognized by host RNA polymerase. These genes encode transcriptional regulators, like Cro, CI, and CII that are involved in the lysis-lysogeny decision as well as the anti-terminator N required for the expression of the delayed early genes. Gene expression proceeds to the delayed early genes only once the lytic pathway has been established. The delayed early genes code for proteins O and P, which are involved in replication of the phage genome, along with the anti-terminator Q. Q acts as a positive control for transcription of the late genes and derepresses the pR' promoter [89]. Q mediated late gene expression begins at ~8 min after infection and initiates four parallel biosynthetic pathways. Three of these pathways are involved in morphogenesis of phage progeny, one each dedicated to producing filled phage heads, tails and tail fibers respectively. The fourth pathway deals with the production of lysis proteins. Each of these pathways is independent and mutually exclusive of the each other. Eventually, the infection cycle concludes with lysis and release of newly formed phage progeny into the extracellular space.

Phage lysis in Gram-negatives

~10^{29} phage infection cycles are estimated to occur per day, making bacteriophage lysis the most common cytocidal event in the biosphere [2]. An estimated 0.37-0.65 gigatons of carbon per year is released in the form of labile organic material from lysis of deep-sea

prokaryotic populations [15]. Each lytic event results in the release of a certain number of virions, which is referred to as the burst size of the phage infection. Evolution puts a constant pressure on phages to fine-tune the lysis timing to optimize progeny release [90]. Change in lysis timing alters the morphogenesis period and thus dictates the number of progeny released [91, 92]. An earlier onset of lysis, which would result in fewer progeny, would be advantageous in a host-dense environment. On the other hand, delaying the onset of lysis would be a better strategy in a host-depleted environment, as it would produce more progeny in one infection cycle. Some phages like T4 even incorporate specialized lysis inhibition mechanisms, which depend on environmental signals to change the lysis time [93-95]. Apart from killing bacteria and phage production, lysis is also associated with production and release of bacterial toxins, such as in the case of Shiga toxin-carrying stx phages [17, 96, 97]. Furthermore, lysis in the case of transducing phages like P1 can generate vectors for transfer of different genetic markers such as antibiotic resistance, pathogenesis factors etc. [98-100]. The omnipresence of phage lysis in all environmental quarters and its role in bacterial pathogenesis renders it worthy of detailed and exhaustive investigation.

Lysis paradigms

The lysis mechanisms of *Caudovirales* infecting Gram-negative bacteria have been studied extensively for decades [101, 102]. These phages encode genes for lysis proteins dedicated to disrupting each component of the cell envelope (Fig. 1.3). Holins permeabilize the IM, endolysins degrade the PG and spanins disrupt the OM. All the lysis systems investigated till date can be broadly classified into two lysis paradigms based on the subcellular localization of the endolysin: the classic/canonical holin-soluble endolysin lysis paradigm and the pinholin-SAR

endolysin paradigm. In the canonical version of phage lysis (Fig. 1.4) represented by phage λ, the endolysin encoded by gene *R* (Fig. 2a), accumulates in the cytosol as fully-folded, active enzyme [103]. Access to PG is controlled by the holin, encoded by gene *S*. The holin gene product S105 accumulates harmlessly in the host membrane until, suddenly "triggering" after reaching a critical concentration [104]. The time of triggering is programmed into the primary structure of S105 and can be affected by even single missense changes [105]. Triggering is correlated with a sudden redistribution of mobile S105 molecules in the IM to two-dimensional aggregates, referred to as "rafts". These rafts cause a collapse in the proton motive force (pmf) and form lethal membrane lesions in the IM in the form of micron-scale holes, [104, 106]. The triggering phenomenon is a universal feature of all holins and can be prematurely induced by energy poisons [107]. With an average size of >300 nm, these large holes allow the endolysin to escape the cytoplasm and attack the PG [108].

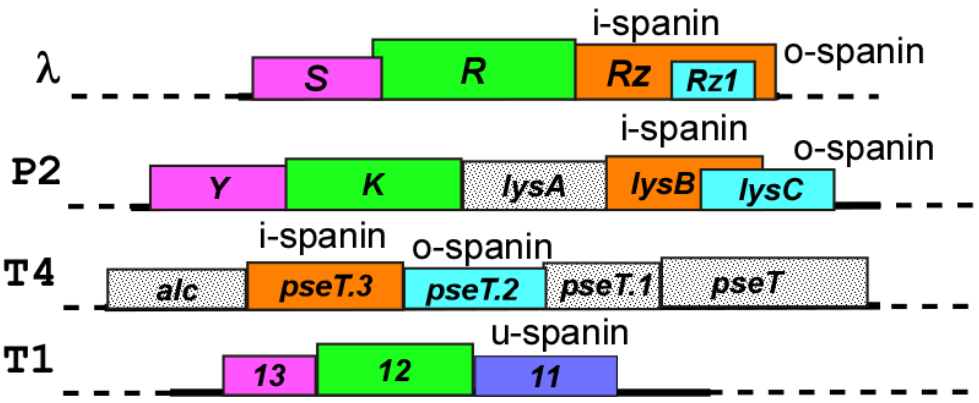


Figure 1.3 Lysis cassette arrangements from the phages λ, P2, T4 and T1. The boxes color coded by pink, green, orange, cyan and purple represent the holin, endolysin, i-spanin, o-spanin and u-spanin respectively in each phage, while the dotted boxes represent genes of unrelated or unknown function. All genes are drawn to scale and labeled with appropriate gene names. The spanin genes are also labeled with their spanin type on top of the gene in each case to denote the differences in architecture.

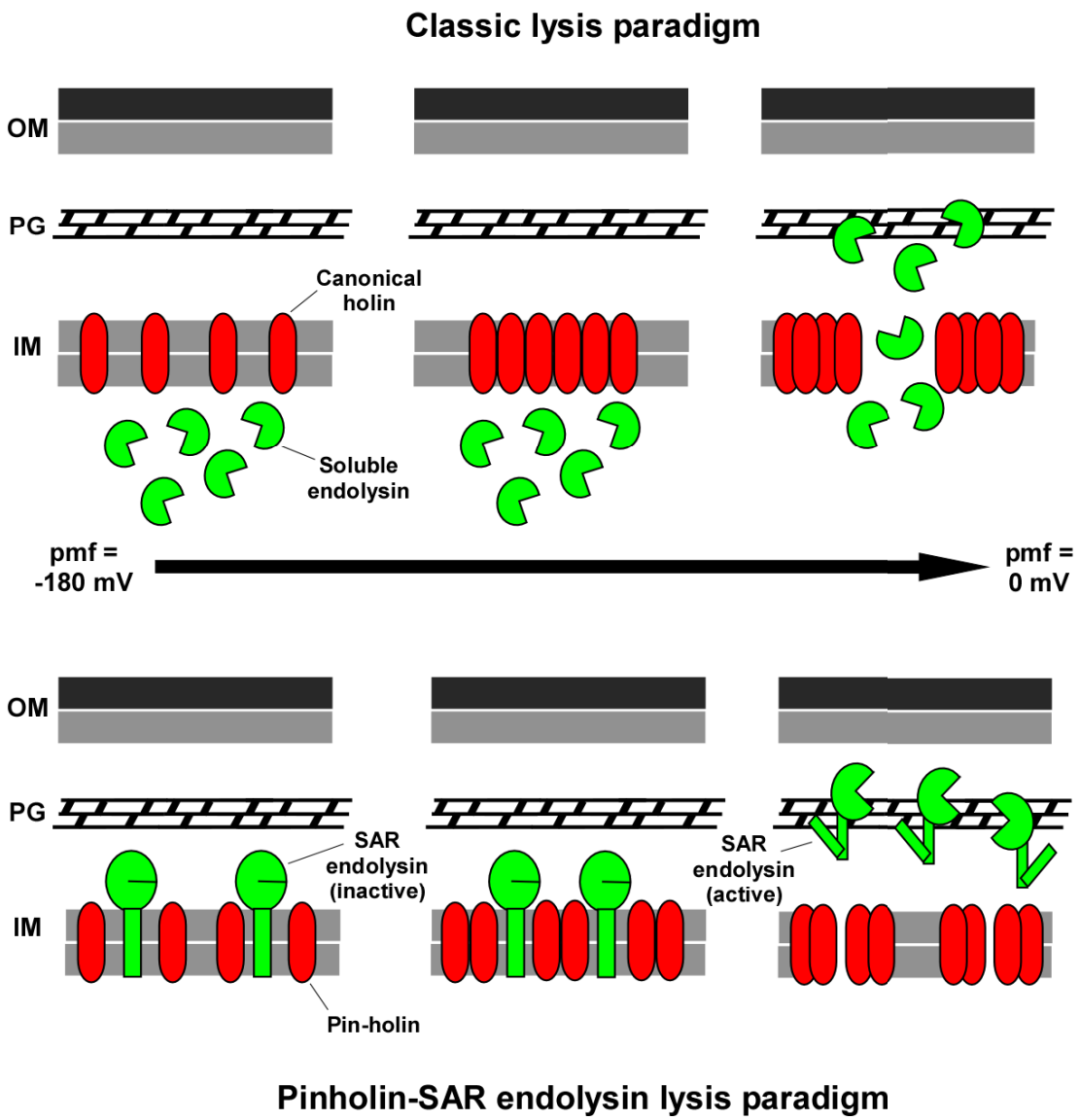


Figure 1.4 Cartoon representation of the classic and "Pinholin-SAR endolysin" lysis paradigms. The cartoon series begins with initial late gene expression on the left and progresses right towards lysis as shown by the arrow. In the canonical lysis paradigm (above), the holin (red ovals) accumulates in the IM while the soluble and active endolysin (open green symbols) accumulates in the cytoplasm. Upon triggering, the holin forms micron scale holes, and allows the cytoplasmic endolysin to degrade the PG. In the "Pinholin-SAR endolysin" lysis paradigm (below), the SAR endolysins (closed green symbols) accumulate in the IM in an inactive form, anchored by a weakly hydrophobic TMD (green stub) alongside the pinholins (red ovals). The pinholins trigger to form nanometer scale holes, good enough to deplete the pmf from -180 mV to 0 mV. This releases the SAR domain, the endolysin gets activated and degrades the PG.

The second pathway to PG degradation, designated as the pinholin-SAR endolysin paradigm (Fig. 1.4), has been investigated using the lambdoid phage 21 as a model [102]. Instead of being cytoplasmic, the endolysin R^{21} accumulates in an inactive form in the periplasm tethered to the IM by an N-terminal SAR (Signal Anchor Release) domain [109]. The SAR label is used to highlight its unique dynamic membrane topology in that the TMD requires the host pmf to stay tethered to the membrane. The control over topological dynamics of R^{21} is exerted by the pinholin $S^{21}68$, the gene product of the λ^{21} holin gene *S* [110]. Like S105, $S^{21}68$ accumulates harmlessly as a homodimer uniformly distributed in the IM, until triggering [111]. Unlike the canonical holins, pinholins form ~10^3 small (~2 nm) heptameric "pinholes", which causes depolarization of the bilayer. Once the host membrane becomes de-energized, the SAR domain exits the bilayer, resulting in re-folding of the endolysin to an enzymatically active form, which then attacks the PG [112, 113].

Not until long ago it was thought that the combined function of holin and endolysin was necessary and sufficient for lysis. However, it has recently been shown that disruption of the OM is also required, as the last step in Gram-negative host lysis. This step is achieved by a class of lysis proteins called spanins [114]. A common feature in most of either of the lysis schemes described above is the presence of a two-component spanin system (Fig. 1.3), consisting of an integral IM protein component, i-spanin and an OM lipoprotein component, o-spanin, which interact to span the entire periplasm and function to subvert the OM [115, 116]. However, adding one more dimension to the lysis paradigms, some phages, including the well-studied coliphage T1, make use of single component spanins [116]. These "unimolecular spanins", designated as u-spanins, are OM lipoproteins that have a C-terminal transmembrane domain (TMD) traversing the IM, thus covalently interconnect both membranes, in contrast to the 2-component spanins,

which form a non-covalent bridge. While the role of holins and endolysins in lysis has been long established, the idea of the spanins being involved in the last step of lysis, breaking the OM barrier, has been established only recently [114, 117]

A detailed review of lysis by Lambda

The lysis paradigm of λ has been studied for more than 40 years, chiefly by our lab, using biochemical and genetic means [101, 102]. The four λ lysis genes S, R, Rz, and Rz1 are tightly linked on the phage genome and are transcribed from the late promoter, pR' (Fig 1.5a). During lysogenic growth, transcription from pR' is terminated by terminators immediately downstream. Upon induction of the lytic cycle, the anti-terminator Q forms a complex with RNA polymerase, resulting in a read-through of the terminators [118]. Most of our experiments involve the lysogen version of *λcI857*, which encodes a temperature-sensitive version of the lysogenic repressor CI [119]. It can be induced for synchronous lysis by shifting growth from 30 °C to 42 °C; i.e., all the prophage population in a growing culture enter the vegetative cycle at the same time. Late gene expression from the pR' promoter starts ~8 min after induction under these conditions. Much of the work described in the following chapters involves either recombining alleles of lysis genes being tested into the prophage chromosome or the use of the medium copy plasmid pRE, carrying the alleles downstream of pR'. Q is supplied by induced λ lysogens or from a compatible vector under IPTG control (pQ_{kan}) [106]. The pRE plasmid system thus provides a platform to replicate *in vivo* expression levels. Our current knowledge of each of the λ lysis proteins is discussed below.

Holin

The lambda holin gene *S* has a dual start motif [120, 121](Fig 1.5a). As a result, *S* encodes two proteins, the holin S105, and the antiholin S107. The translational start for S107 is two codons upstream of S105 [91]. Though they share the same coding sequence except for the two additional residues, they have opposing functions. The lysis timing of λ is programmed into the holin primary structure [92]and single missense changes within the holin gene, even conservative, result in dramatically accelerated or delayed lysis [106]. The holin gene product S105 is a 105 aa IM protein with three transmembrane domains in an N-out, C-in topology [122] (Fig 1.5b). S105 accumulates harmlessly across the IM as mobile homodimers without affecting the membrane integrity of the IM [106]. After reaching a critical concentration, which for wild-type S105 would be at ~50 min after induction of lysis, the holin undergoes "triggering" [104]. According to the "death raft" model, triggering reflects the sudden rearrangement of the holins from uniformly distributed mobile dimers to punctate and immobile 2D aggregates known as "rafts". These rafts are characterized by close helical packing of TMDs that excludes lipid (Fig 1.5c). Any changes that would affect the helical packaging would in turn alter the lysis timing. At a critical concentration, the rafts lose the ability to support the pmf. At this point, holin molecules undergo quaternary conformational changes initiate an opening in the raft, eventually leading to hole formation. Cryo-EM and tomography studies of cells that had undergone S105 triggering showed micron scale interruptions in the cytoplasmic membrane [107, 108]. Recent cysteine-scanning accessibility studies showed that all of the S105 proteins participate in the holes with the TMD1 and TMD3 lining the boundary of the hole (Fig 1.5c) [123].

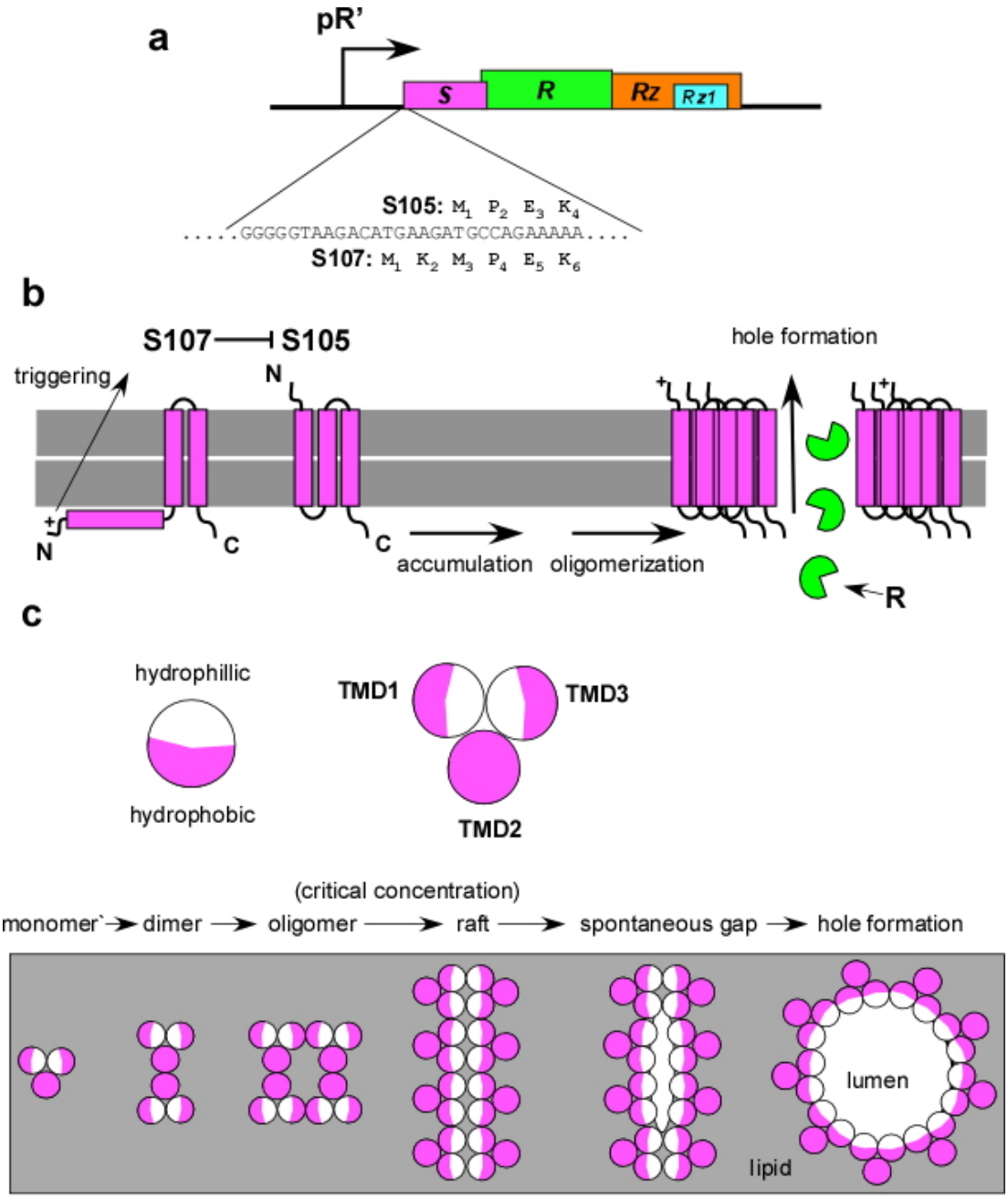


Figure 1.5 The lambda holin. a) The lysis cassette of lambda and the dual start motif of the *S* gene is shown. b) The topology of S107 and S105. In this configuration, S107 negatively regulates S105. At the time of triggering the S107 TMD1 flips its N-terminus into the periplasm. c) Top-down view of the holin TMDs. The hydrophobic and hydrophilic faces are shown. Below the lipid bilayer is shown in grey. The death raft model for oligomerization and rearrangement of the holin molecule is shown. The holin accumulates as dimers. At a critical concentration, the holin forms rafts, which get large enough to disrupt energy production of the cell. When pmf is lost, the holin rearranges to form a hole. The hydrophilic faces of TMD1 and 3 face the lumen.

The upstream translational start of the antiholin S107 adds a positively charged residue at the N-terminus, preventing the TMD1 from assuming the N-out topology [91]. As a result, S107 has an N-in, C-in topology (Fig 1.5b). Disruption of the dual start motif revealed that S105 is still lytic in the absence of S107, but the converse is not true [124]. The inhibitory capacity of the antiholin is dependent on its additional N-terminal positive charge. However, energy poisons can eliminate the inhibitory effect of the antiholin and convert S107 into a lysis effector. At physiological levels, the production ratio of S105 to S107 molecules in vivo is 2:1; this ratio, determined by the secondary structure of the *S* transcript, dictates lysis timing [91]. The role of the antiholin is to inhibit S105 by forming inactive heterodimers [125]. Once a critical concentration of S105 molecules is reached and the pmf is disrupted, TMD1 of S107 flips into the periplasm [121]. This turns S107 from an inhibitory molecule to a functional holin, effectively tripling the pool of active holins involved in triggering, based on the 2:1 S105:S107 ratio [91].

The holin being the only determining factor of the onset of lysis has allowed phages to evolve to optimize the lysis timing [90, 92]. A short latent period would result in fewer progeny but would be beneficial in high host density environment. In a low host density scenario, a lengthier infection cycle would be advantageous as it would increase the burst size. The allele-specific differences in lysis timing of S105 most likely reflect the malleability of the temporal schedule imposed by holins [105].

Endolysin

Endolysins are muralytic enzymes that account for the second step of lysis, the degradation of the PG. The λ endolysin encoded by the gene *R* accumulates in the cytoplasm in a

fully active form. All known endolysins can be categorized into amidases, endopeptidases, glycosylases (also known as lysozymes) and transglycosylases, depending on their mechanism of targeting linkages in the PG meshwork [126]. While the amidases and endopeptidases break oligopeptide crosslinks, the other two target glycosidic linkages. The λ endolysin belongs to the transglycosylase family and cleaves the ß-1,4 glycosidic linkage between MurNAc and GlcNAc. A C-terminal fusion of the R endolysin with a full-length beta-galactosidase domain, which forms a tetramer >480 kDa in size, was also shown to be lytically functional [103]. Despite the diversity in mechanisms, complementation studies have demonstrated that endolysins will function lytically with any holin that can provide access to the PG, indicating that there is no specific interaction between holins and endolysins [127].

Spanins

The lambda spanins *Rz* and *Rz1* share an embedded genetic architecture (Fig 1.3) [128]. While embedded genes are not uncommon in prokaryotic biology, *Rz* and *Rz1* are unique in the sense that they share the same DNA sequence and encode proteins involved in the same function. Rz is the prototypical i-spanin whereas Rz1 is the prototypical o-spanin [129]. Absence of either Rz or Rz1 results in a terminal spherical cell phenotype [114]. This phenotype is a result of conversion of cells from rod to sphere shape after the PG is degraded by the endolysin. These spheres, which are essentially are the cytoplasmic contents bound by the OM, are fragile and can be broken down by the shearing forces in aerated cultures. This was the reason that spanins and OM disruption were overlooked for more than two decades. The spanin lysis defect is evident only in media supplemented with divalent cations or in non-shaking conditions. The divalent cations are believed to stabilize the OM by forming ionic bridges with phosphate groups of the

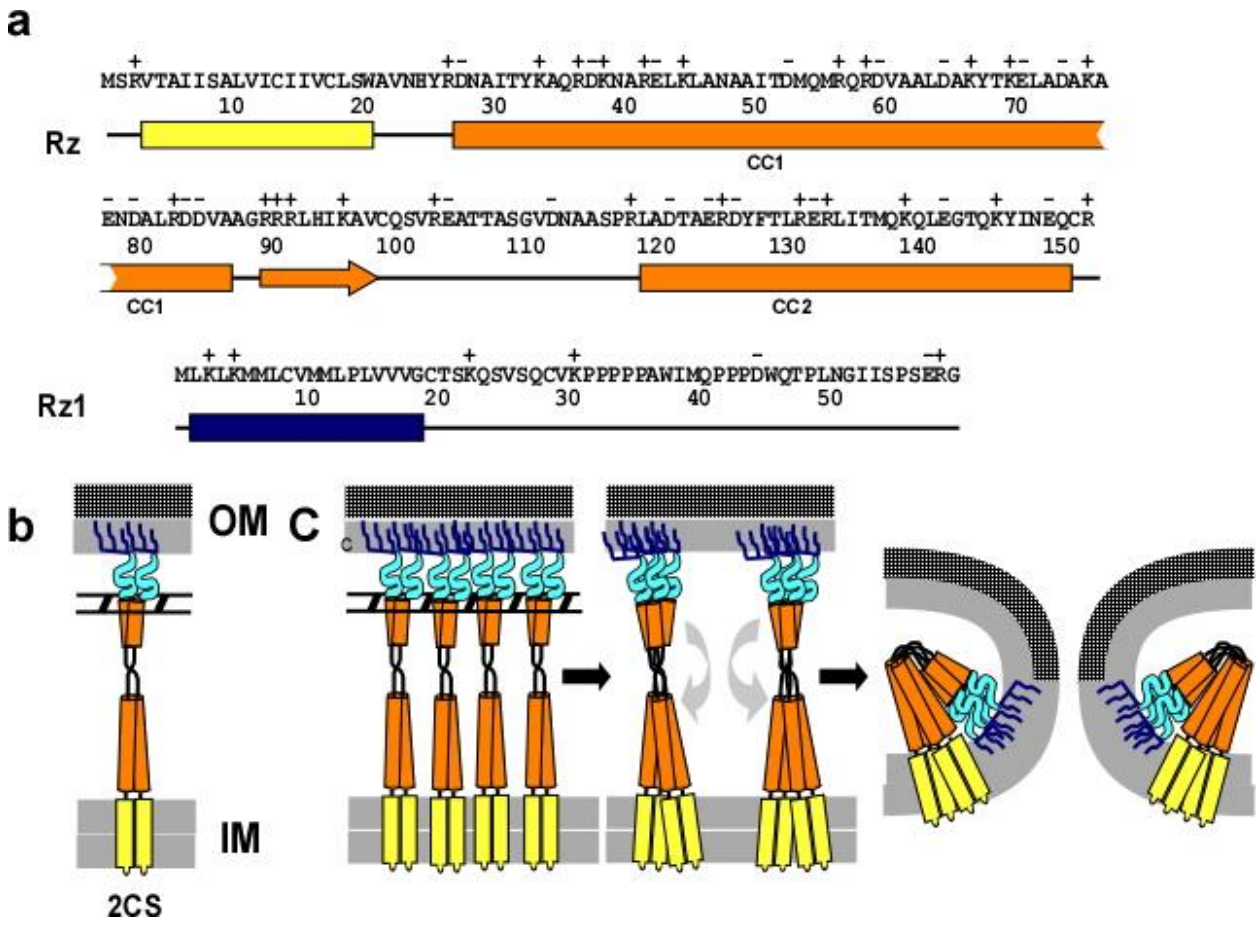


Figure 1.6 The lambda spanins. a) Predicted secondary structures of lambda i-spanin Rz and o-spanin Rz1. The TMD in Rz is shown as yellow rectangle while the lipoylation signal sequence in Rz1 is shown as dark blue rectangle. Orange rectangles and arrows indicate predicted alpha helical and beta sheet domains respectively in Rz. The predicted coiled-coil domains CC1 and CC2 and the flexible linker region connecting them are shown. b) Cartoon representation of the Rz Rz1 spanin complex. The i-spanin Rz is embedded to the IM by an N-terminal TMD (yellow rectangle) and has a periplasmic domain that constitutes two alpha helices (orange cylinders) connected by a linker, predicted to form coiled coils. The o-spanin Rz1 is attached to the inner leaflet of the OM via three fatty acyl chains (dark blue lines) and has a periplasmic domain (cyan) predicted to be unstructured. The i- and o- spanins interact through their C-termini to form the spanin complex, linking the IM and OM through the PG meshwork. c) Cartoon representation of the model for the Rz Rz1 function. The Rz Rz1 spanin complexes accumulate within the constraints of the PG meshwork, connecting the IM and OM (left panels). Once the PG is removed by the endolysin, these complexes are free to undergo lateral diffusion and conformational changes (middle panels). These events bring both the membranes together, leading to fusion of the IM and OM (right panels) and release of the phage progeny through the fusion pore.

LPS molecules on the outer leaflet of the OM. These spherical cells, however, are physiologically dead due to permeabilization of the IM by the holin.

The *Rz* gene was discovered in 1979 through transposition mutagenesis of the λ chromosome [130]. Insertion of the transposon into the *Rz* locus rendered the lysogen deficient for lysis in magnesium-supplemented media. The gene product of *Rz* is a type II integral IM protein, 153 aa long, with an N-terminal TMD and a highly charged C-terminal domain that extends into the periplasm (Fig 1.6a). The periplasmic domain of Rz is comprised of two predicted alpha helices that are predicted to form a coiled-coil structure (Fig 1.6b). The N-terminal TMD anchor to the IM was shown to be necessary for Rz function [115]. However, the TMD of Rz could be replaced with that of FtsI without any effect on function. This suggested that the role of the TMD is only to physically tether Rz to the IM. Rz accumulates as a homodimer linked by two intermolecular disulfide bonds between periplasmic cysteines at position 99 and 152 [129] (Fig 1.6b). However, non-covalent dimers are formed in vitro, even in the absence of the cysteines.

The *Rz1* gene was first identified as an unexpected 6.5 kDa protein in experiments meant to express *Rz* [131]. It was demonstrated that this protein was encoded by a gene in +1 reading frame within *Rz*, which was named *Rz1*. Further biochemical studies confirmed that Rz1 was an OM lipoprotein [132]. Rz1 is translated as a 60 aa prolipoprotein, which gets processed into a 40 aa mature lipoprotein (Fig 1.6a). The presence of Thr and Ser in +2 and +3 positions targets mature Rz1 to the OM [133]. Translocation of Rz1 to the OM is carried out by the Lol system as described above. Owing to its high proline content (10 of 40 aa in the mature Rz1), Rz1 is predicted to be mostly unstructured (Fig 1.6b). Like Rz, Rz1 also accumulates as a homodimer,

linked by an intermolecular disulfide bond between the periplasmic cysteines at position 29. Unlike Rz, in vitro dimer formation is dependent on Cys29 [129].

Multiple sources of evidence indicate Rz and Rz1 interact through their C-termini to form a spanin complex that physically links the IM and OM [115]. First, Rz could be co-immunoprecipitated along with Rz1-his in a pull-down assay. Also, expression of Rz in the absence of Rz1 results in an unstable Rz product. This suggests that there is a protease-sensitive site near the C-terminus of Rz that gets exposed in the absence of Rz1. Furthermore, fluorescence studies using an N-terminal GFP translational fusion of Rz revealed that GFP-Rz exhibits a punctate distribution when co-expressed with Rz1 [114]. When Rz1 was absent, GFP-Rz was uniformly distributed throughout the IM, suggesting that the interaction of Rz with Rz1 across the width of the periplasm was responsible for the punctate foci phenotype. Notably, the spanin complex forms even before the PG has been removed by the endolysin. As both Rz and Rz1 accumulate as dimers, the spanin complex exists as a heterotetramer linked by three intermolecular disulfide bonds at Cys99 and Cys152 in Rz and Cys29 in Rz1(Fig 1.6c) [129]. Substitution analysis with different combinations of Cys to Ser changes established that at least one intermolecular disulfide bond at either Cys152 of Rz or Cys29 of Rz1 was necessary and sufficient for spanin function. In addition, it was also determined that host Dsb system, heretofore shown to be involved exclusively in intramolecular disulfide bond formation, was required for formation of these intermolecular linkages.

The embedded architecture of *Rz Rz1* poses a problem for any genetic/mutational analysis, as changes introduced into one gene within the shared DNA could also alter the coding sequence of the other gene. To address this, a synthetically separated spanin pair was constructed with the *Rz1* gene downstream of *Rz* [134]. The separated architecture allowed for the

manipulation of one gene without disturbing the coding sequence of the other. Moreover, to prevent the separated genes from undergoing recombination within the shared sequence, the nucleotide sequences of *Rz* and *Rz1* were altered to be distinct without introducing any changes into their amino acid sequences. Phages carrying these separated spanin genes were indistinguishable from the parental phages in terms of the rate of lysis, plaque size, and protein levels. In addition, genetic studies with the separated pair of *Rz Rz1* provided insights about the functional significance of different domains in the spanin complex [134]. Several inactivating mutations in the coiled-coil domains CC1 and CC2 of Rz were isolated. On the other hand, the linker domain between CC1 and CC2 was shown to be flexible and genetically silent, allowing insertions of Gly-Ser oligopeptide repeats. Most of the non-functional mutations in Rz1 mapped to the proline-rich region (PRR), especially the central penta-proline stretch. It was also confirmed through Gly-Ser repeat additions that, similar to the flexible linker in Rz, there exists a flexible linker region between the lipoylated Cys and the PRR in the N-terminus region of mature Rz1. Complex formation, as assessed by coimmunoprecipitation, was not affected for most of the non-functional Rz alleles, except RzE150G. However, an R59E change in Rz1 suppressed both the lysis and spanin complex-formation defects. Additionally, the combination of RzE150R and Rz1R59E alleles produced a functional spanin complex, implying a potential salt bridge interaction at these sites. In another suppressor study, mutations in the juxtamembrane region of CC1 in Rz were isolated that could suppress non-functional mutations of both Rz as well as Rz1 [135]. The suppressor mutations were mostly polar insertions into the hydrophobic core of CC1. It was proposed that they might affect the stability of the juxtamembrane region, leading to a conformational state that overcomes the fusion block created by the primary mutation.

In contrast to the detailed picture emerging from the biochemical and genetic studies, there is very little structural information for Rz-Rz1. Consistent with structural predictions, circular dichroism analysis indicated that sRz (soluble periplasmic domain of Rz) exhibits alpha-helical content whereas as sRz1 (soluble periplasmic domain of Rz1) is primarily unstructured [52]. Mixing of sRz and sRz1 forms a complex of equimolar stoichiometry accompanied by a significant increase in the alpha-helical content. This indicates potential rearrangement/elongation of secondary structures triggered by the C-terminal interaction of Rz and Rz1. TEM single particle analysis of the soluble domains mixed together revealed supramolecular rod-shaped structures that were measured to be ~24 nm long. This measured length of the spanin complex is in agreement with our cryo-EM measurements of the periplasmic width under the standard growth conditions used in our lab.

Based on similarities to known class I viral fusion systems, it was proposed that λ spanins disrupt the OM by fusing it with the IM [115] (Fig 1.6c). In this model, spanin complexes accumulate in both the membranes in a metastable membrane-spanning conformation. The spanin complexes are constrained to the lacunae of the PG meshwork through which they are threaded. Once the PG is degraded by the phage endolysin, the spanin complexes undergo a conformational change bringing both the membranes together, eventually fusing the IM and OM. Thus, spanins mediate a topological solution instead of a degradative one, for the last step of lysis. The membrane fusion model was supported by recent data that demonstrated spanins can fuse *E. coli* spheroplasts [117]. Spheroplasts are produced by slowly treating *E. coli* cells with EDTA and lysozyme in the presence of 20% sucrose [136]. The EDTA destabilizes the LPS layer on the OM, subjecting the PG to lysozyme degradation, resulting in a spherical cell bound by the IM and osmotically stabilized by the sucrose. The basis of this spheroplast fusion assay

was to replicate the spanin complex in two lipid bilayers, both the bilayers being IM in this case [117]. For this to be possible, Rz and Rz1 would need to be anchored to the IM with the periplasmic domains exposed, so that they can interact with each other in neighboring spheroplasts. This required using the missense allele of Rz1 (denoted *imRz1*), in which the two residues at +2 and +3 positions are substituted with Asp. The substitution abolishes the OM lipoprotein sorting signal, thus anchoring Rz1 to the outer leaflet of the IM. GFP and mCherry were also co-expressed to fluorescently mark the spheroplasts carrying Rz and Rz1 respectively. Spheroplasts expressing Rz were shown to undergo efficient fusion with spheroplasts expressing imRz1. In contrast, lysis-defective missense alleles of either Rz or Rz1 showed an increase in the number of adhered spheroplasts but did not support fusion. While these results support the idea of spanins being fusogenic, whether IM-OM fusion occurs as part of the normal lysis pathways still needs to be verified.

Membrane fusion

Since we propose that spanins achieve OM disruption through the fusion of the inner and outer membranes, a detailed generic review of membrane fusion is provided here. Membrane fusion is key to several fundamental biological processes with diverse functions [137]. Membrane fusion is the process of merging two separate lipid membranes into a single continuous bilayer and can occur between different cells, between different or similar intracellular compartments, between plasma membrane and membrane-bounded structures like viral particles or vesicles, and between cellular membranes [138]. Entry and exit of enveloped viruses, vesicle tracking, exocytosis and endocytosis, the formation of syncytia of muscle cells,

the fusion of haploid eukaryotic cells during mating, and now phage lysis as proposed recently are some of the varied examples that are dependent on membrane fusion [139-142].

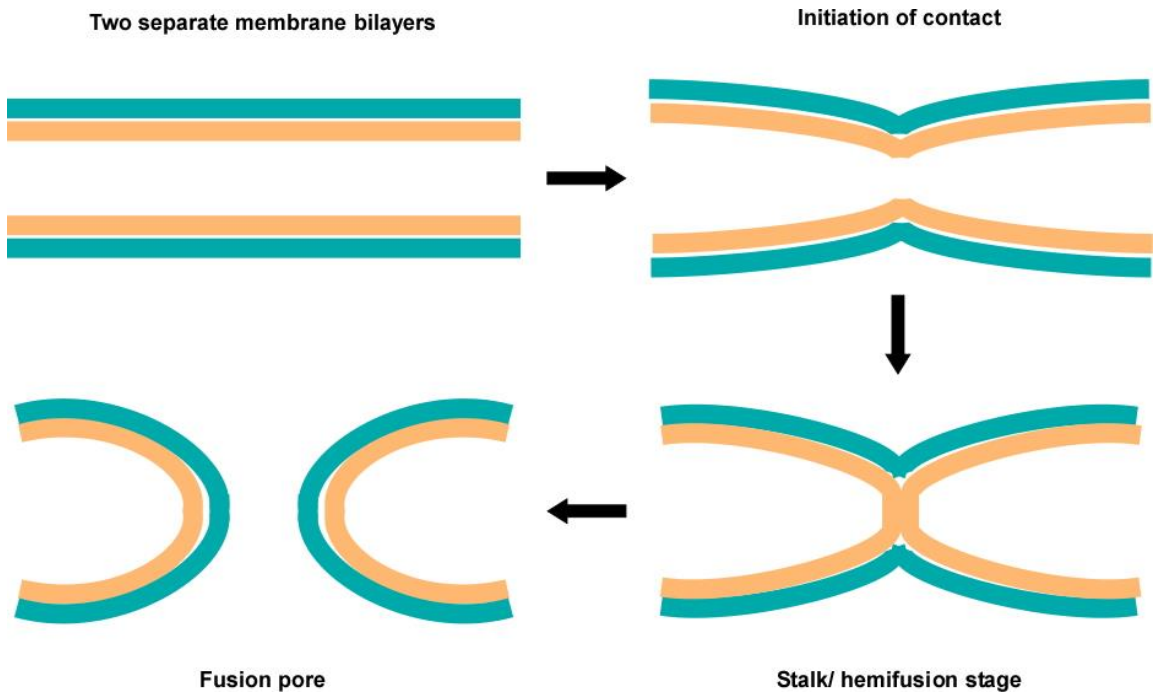


Figure 1.7 Stalk model of membrane fusion. Different steps in the stalk model are shown in a clockwise manner (I) Two separate membrane bilayers. (II) Initiation of membrane contact between two separate membrane bilayers. (III) Formation of a fusion stalk – the hemifusion state. (IV) Establishment of a fusion pore. The proteins involved in the fusion process are not shown for simplicity.

Despite the diversity in fusion proteins, all membrane fusion reactions follow the same elementary steps of the proposed "stalk model", shown in Fig. 1.7: a) bringing two membranes within proximity to each other; b) merging of membranes, also referred to as the intermediate state or the hemifusion state; and c) formation/opening of an aqueous fusion pore. However, under biologically relevant conditions, apposed bilayers do not fuse spontaneously [143, 144]. Initiation of fusion requires localized destabilization of the membrane bilayers [145, 146]. Owing to the high energy cost associated with the destabilization, the initiation step is the rate limiting-

step in the fusion process [147]. However, this energy barrier can be lowered by fusion proteins that are ectopically localized to the membrane bilayer and bend the membranes by pulling them together [148, 149]. This is reflected in the hemifusion step, where the outer leaflets of the donor and target membrane are highly curved and fused while the inner leaflet is intact, forming a stalk-like structure. Subsequently, this stalk radially expands until the trans-monolayers meet to form a new bilayer called hemifusion diaphragm, the rupture of which forms the fusion pore, completing the fusion process [145, 146].

There is also an alternate pathway designated as “direct fusion”, although with less supporting evidence [150]. According to this scheme, fusion proceeds via the formation of a form a “hemichannel” structure by fusion proteins lined up on both sides of the opposed membranes. A fusion pore, which is established once the two hemichannels come together, allows the exchange of lipids and aqueous contents between the two membranes. Structural studies of fusion proteins from yeast, plants, and mammals indicate that even though they are highly complex and diverse, the membrane fusion pathway is universal [151, 152]. It is the regulatory mechanisms that determine the differences in these fusion systems.

Here, we will review two different types of viral entry proteins that have the strongest similarities to spanins. Class I viral membrane fusion proteins contain alpha helical coiled-coils, whereas class II viral membrane fusion proteins consist of beta-sheet secondary structures [153]. Both the classes of viral entry membrane fusion proteins need to be triggered by environmental factors, like pH change, or protein-protein contact, like receptor binding, to initiate the fusion cascade (Fig 1.8). Triggering starts a series of events that lead to changes in fusion protein oligomeric state and conformation, eventually ending in a stable postfusion confirmation [154]. During triggering, a hydrophobic fusion peptide is first exposed, which re-orients to interact with

the outer leaflet of the target membrane. These fusion peptides can be inserted either in their trimer form or as monomers, which trimerize after insertion [155-157]. The extended state of the fusion proteins after insertion into the target membrane is known as the "pre-hairpin" conformation. The pre-hairpin structure then folds into a full-hairpin conformation, which brings the target membrane and virus membranes into proximity. The insertion of fusion peptides also exerts a force on the outer leaflets which causes distortions in the opposing membranes, bending the membranes towards each other [158]. Next, a "hemifusion intermediate" or the "lipid stalk" is created by the folding back of the hairpin [144]. Finally, a fusion pore forms within the stalk allowing the mixing of the viral contents with the host cytoplasm. The expansion of the pore is believed to be the most energetically-demanding step [159], and likely involves several fusion proteins [142, 153]. While prefusion structures have some degree of structural variability, all postfusion structures are trimeric hairpins. Specific examples of Class I and II viral fusion proteins are discussed below.

Class I Viral Fusion Protein

The human influenza virus hemagglutinin (HA) is a prototypical Class I viral fusion protein with information available on pre-fusion as well as postfusion structures. [154]. Like other, class I viral fusion proteins, HA is synthesized as inactive precursors that rearrange into trimers. The inactive precursor form of HA is called HA0, which gets processed into an N-terminal HA1 domain and a C-terminal HA2 domain, linked through a disulfide bond [160].

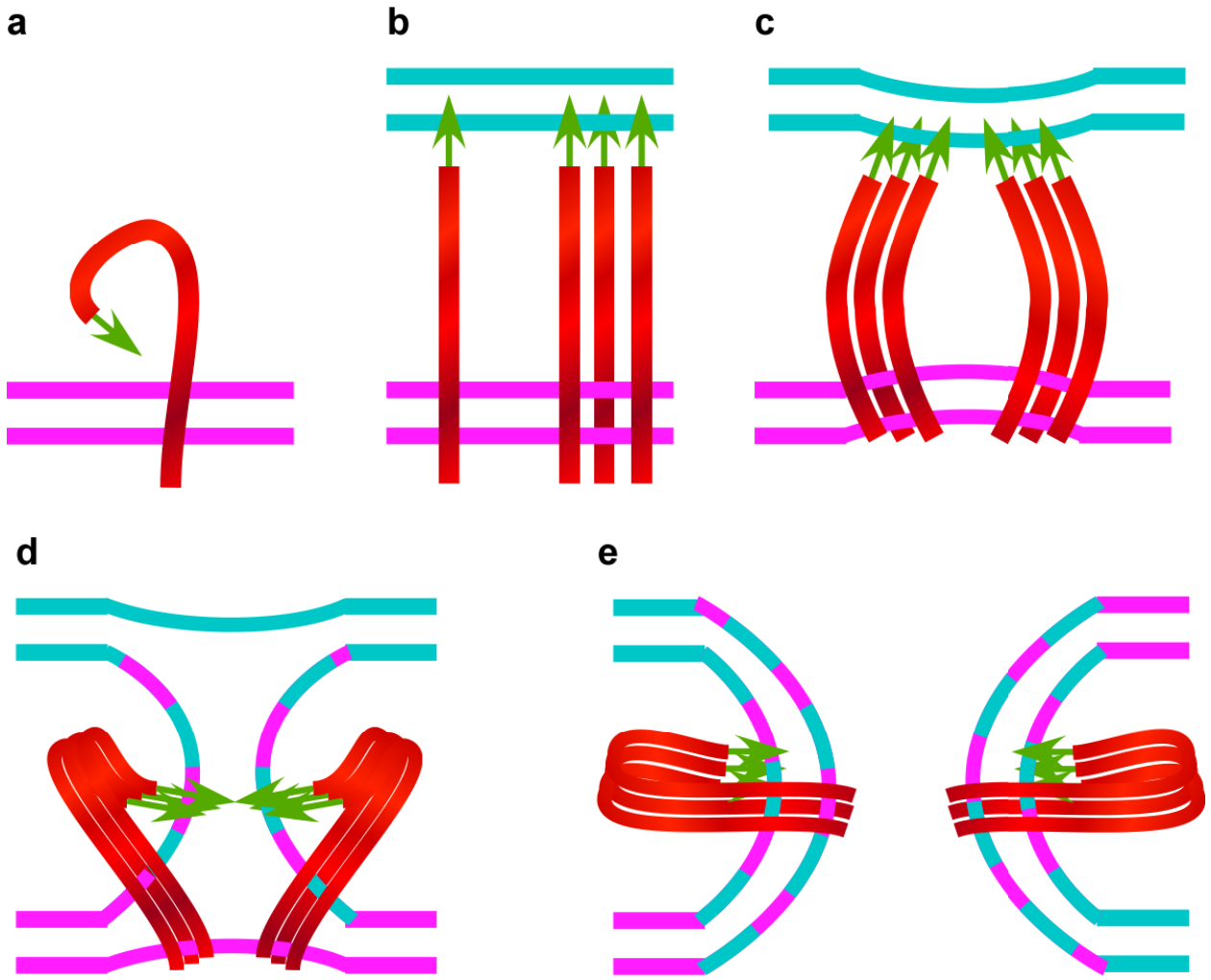


Figure 1.8 General model for viral membrane fusion proteins. (a) A metastable viral fusion protein in the prefusion conformation (red) has its fusion peptide (green arrow) hidden or inactive. (b) The fusion protein is triggered, and the fusion peptide or loop inserts into the target cell membrane (blue) to form a monomeric or trimeric extended intermediate that bridges the two membranes. (c) Protein insertion and refolding cause initial distortions in the viral (purple) and/or target membranes. (d) Fold-back of the proteins in a hairpin-like fashion promotes the formation of a hemifusion intermediate in which the contacting membrane leaflets mix. (e) The final packing of the outer layer onto the core trimer transitions the protein to the postfusion hairpin and drives the opening of the fusion pore.

HA1 contains the receptor binding site, which recognizes sialic acid residues on glycoproteins and glycolipids [156]. With the globular HA1 “head” at the top and the HA2 “stalk” connecting to the TMD, the HA forms a vertically oriented trimer structure on the surface of the virus. The

HA2 stalk is composed of a three stranded alpha helical coiled-coil, one from each monomer. There is a ~20-25 aa fusion peptide at the N terminus of HA2, which is freed by proteolytic processing of the HA0 form. Post cleavage, the fusion peptide is still buried in a pocket within the trimer interface of the stalk. Once inside an endosome, the pH change causes the HA1 heads to separate. This allows protonation of charged residues within the pocket, which releases the fusion peptide from the trimer interface [161]. The fusion peptide gets extended about 100 Å toward the top of the molecule and interacts with the target membrane. Through multiple conformational changes, the C-terminal region of HA2 is packed into the grooves of the coiled coil. This essentially forms the outer layer of the postfusion hairpin and brings the fusion peptide and TM domains to the same end of the trimer. Several HA trimers, inserted into the target membrane, are required to mediate fusion. There is also evidence that suggests that trimers that are not inside the contact region play a role in the expansion of the fusion pore [162].

Class II viral membrane fusion proteins

The alphavirus E1 protein is one of the best examples of the class II fusion proteins [154]. These proteins are all initially synthesized as part of a polyprotein that also contains an N-terminally located companion or chaperone protein [163]. Structural studies indicate that alphavirus E1 exists in a prefusion monomer form [164] and a postfusion trimer form [165]. In the monomeric form, alphavirus E1 is an elongated molecule with three domains (DI–III) composed primarily of β-sheets. DI is the central domain, connected to DII with a flexible hinge and to DIII on the other end with a linker. The DII domain has a fusion loop at its tip, while DIII has an immunoglobulin (Ig)-like fold. The DIII C-terminus connects to a stem region followed by the TM anchor. The chaperone/companion protein p62 is associated with E1 during its

synthesis within the endoplasmic reticulum (ER). In the late secretory pathway from the ER, p62 gets processed into a mature TM associated E2 domain, which is an elongated molecule with three Ig-like domains and a peripheral E3 peptide. The E3 domain binds to and protects E1 from low pH during exit. After the exit, the extracellular neutral pH causes the release of the E3 domain and completes the maturation of E1 [166]. Now the E1is primed for fusion during endocytic entry. The fusion proteins are organized as "spikes" composed of trimers of E2-E1 heterodimers on the surface of the virus. The E2-E1 heterodimers are tightly but noncovalently associated, with E1 oriented tangentially to the virus membrane and the fusion loop clamped between the Ig-like domains of E2. Upon endocytosis, the low pH environment causes a rearrangement of E2-E1 interactions [167]. The fusion loop at the tip of DII is inserted into the target membrane. The three DI-DII regions associate to form the core trimer while DIII and the stem pack against the core to form a postfusion conformation with the fusion loops and TM domains at the same end of the molecule. Despite structural differences, this trimeric hairpin arrangement is similar to that of postfusion HA.

Similarities between spanins and viral fusion proteins

Given the predicted structural properties and molecular characteristics of spanins, it is likely that the spanins function in a similar manner to other viral fusion proteins. The coiled-coil domains, similar to ones that have been shown to be a key element of class I viral fusion proteins, are well conserved in Rz from phage λ and its functional homologs. On similar lines, gp*11* and its homologs all have predicted beta-sheet secondary structures, analogous to class II viral fusion proteins. The spanin complex in the case of two-component spanins, as well as the unimolecular spanins are believed to make direct contact with the two membrane bilayers, IM

and OM [115, 116], thus imitating the pre-hemifusion extended stage of the viral fusion proteins. Furthermore, biochemical studies with the soluble domains of Rz and Rz1 revealed that they undergo significant conformational changes upon interaction[52]. It can be thought that these conformational changes provide the free energy required for the fusion reaction. Another notable feature is the unusually high number of prolines in the periplasmic domain of Rz1. The role of polyproline stretches in fusion have been demonstrated recently for the p15 protein of non-enveloped reovirus [168]. Altogether, the structural features of spanins are reminiscent of different kinds of viral fusion proteins. Moreover, spanins act in the last step of lysis for phage release. This suggests that there is no requirement to reverse the spanin complex for another round of fusion which is analogous to viral entry or exit by fusion.

Questions to be addressed

Almost all our knowledge about spanin function comes from studies of the two-component spanins, Rz and Rz1 from lambda. Not only is there a lack of experimental information on any other two-component spanins, but also nothing is known about the unimolecular spanins. The number of sequenced bacteriophage genomes deposited in Genbank has increased exponentially over the last decade. A significant fraction of these genomes is from phages of Gram-negative hosts. In comparison, the number of annotated spanins is astonishingly low. In chapter II, we present a spanin database created by identifying potential spanins in ~580 phages of Gram-negative hosts. We also performed a detailed phylogenetic analysis of spanin sequences and grouped them into families. In addition, we analyzed secondary structure predictions, patterns in functional domain organization, and periplasmic cysteine composition. Chapter III addresses the localization and regulation of the unimolecular spanin gp*11*. We used

topological alleles of gp*11* that localize to only one of the membranes to show that both the membrane localization signals are true and necessary for function. Using a C-terminal GFP tagged allele of gp*11*, we document the punctate distribution of gp11 foci, indicating oligomerization of gp11. Notably, this oligomerization occurred even in the presence of PG. Another interesting finding of this chapter was that gp*11* is regulated at multiple levels: first at the post-translational processing level; then by expression from an internal start that results in a short gp*11* stub localized to only IM; and third by the PG which acts as a physical barrier to gp11 interactions and conformational changes. Chapter IV reports a detailed genetic and biochemical analysis of gp*11*. Multiple loss of function mutations spread across the periplasmic domain of gp*11* were isolated by using random PCR mutagenesis. Through crosslinking and time-lapse fluorescence microscopy, it was shown that these mutants vary in oligomerization and accumulation patterns. The highlight of the chapter is the spheroplast fusion assay that showed gp*11* can fuse two differently fluorescently labeled spheroplasts, supporting our idea of membrane fusion as the mode of action for gp*11*.

CHAPTER II

PHAGE SPANINS: DIVERSITY, TOPOLOGICAL DYNAMICS AND GENE CONVERGENCE*

Introduction

It has long been thought that holin-endolysin function was necessary and sufficient to effect bacteriophage lysis and achieve liberation of progeny virions [92, 101, 108, 110] , except in phages of the mycolata, where a third functional class, the Lysin B esterases, are essential to degrade the waxy outer membrane [169, 170]. Recently, however, another functional class of proteins, the spanins, have been shown to be required for disruption of the outer membrane (OM) in Gram-negative hosts [114, 115, 117, 128]. To establish a context for spanin function, a brief overview of the holin-endolysin pathway is required.

Holin-endolysin lysis pathways

In the canonical version of phage lysis (Fig. 2.1a) represented by phage λ, the endolysin, encoded by gene *R* (Fig. 2.1b), accumulates in the cytosol as fully-folded, active enzyme. Access to its substrate, the peptidoglycan (PG) is controlled by the holin, encoded by gene *S*. The holin gene product S105 accumulates harmlessly in the host inner membrane (IM) until, at a time programmed into its primary structure, suddenly "triggering" after reaching a critical concentration [104, 108]. Triggering is correlated with a sudden redistribution of the holin molecules in the IM to two-dimensional aggregates, referred to as "rafts", and a collapse in the

*This chapter has been modified from a manuscript submitted by Kongari et al., to BMC Bioinformatics (May 2018)

proton motive force (pmf). Upon triggering, lethal membrane lesions occur in the IM within these rafts, in the form of micron scale holes, the boundaries of which are lined by the two of the three transmembrane domains (TMDs) of S105 [123]. This breach in the IM allows the endolysins to escape the cytoplasm and attack the PG (Fig. 2.1a).

A second pathway to PG degradation, designated as the pinholin-SAR endolysin paradigm, has recently been described in detail, using the lambdoid phage ϕ21 (Fig. 2.1a) as a model [110]. In this pathway, the muralytic enzyme, R^{21}, is secreted through the host translocon and accumulates in an inactive form in the periplasm tethered to the IM by an N-terminal TMD [110]. The membrane-tethered state of R^{21} requires the host pmf, so when the host membrane becomes de-energized, the TMD exits the bilayer, resulting in re-folding of the endolysin to its enzymatically active form, which then attacks the PG [112]. R^{21} has been designated as a SAR endolysin (Signal Anchor Release) to highlight the dynamic membrane topology. Control of R^{21} is exerted by the pinholin, $S^{21}68$, a product of the ϕ21 *S* gene. Like S105, $S^{21}68$ accumulates harmlessly as a homodimer uniformly distributed in the IM until triggering [104, 111]. Unlike the canonical holins, however, pinholins form ~10^3 small (~2 nm) heptameric "pinholes", which results in the depolarization of the bilayer and activation of the SAR endolysin [110, 112].

Origin and characterization of the prototype spanins

For both lysis pathways, it has recently been shown that the spanins, encoded by the *RzRz1* genes in both lambda and ϕ21, are also required [114]. The lambda *RzRz1* genes had originally attracted attention because of their bizarre architecture, with *Rz1* embedded in the +1 reading frame of *Rz* (Fig. 2.1b). Knockouts of either gene caused an absolute lysis defect, in which the host cells were converted into fragile spherical forms bounded by the intact OM [128].

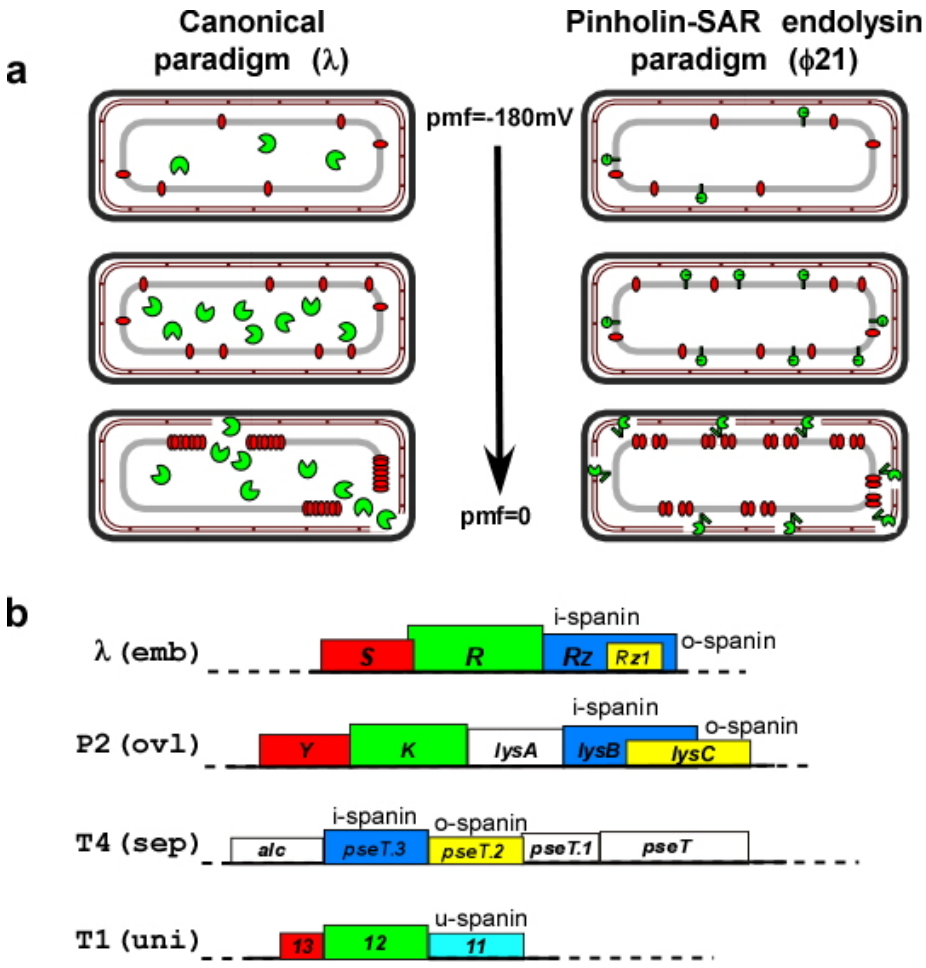


Figure 2.1 Lysis paradigms and cassettes. a) Cartoon representation of the canonical and "Pinholin-SAR endolysin" lysis paradigms of phages λ and ϕ21. The cartoon series begins with initial late gene expression and progresses downwards towards lysis as shown by the arrow. The cellular envelope components IM, PG and OM of a Gram-negative bacterial cell are shown as grey rectangle, hatched brown rectangle and black rectangle, respectively. In the canonical lysis paradigm (left), the holin (red ovals) accumulates in the IM while the active endolysin (open green symbols) accumulates in the cytoplasm. In the "Pinholin-SAR endolysin" lysis paradigm as in the phage ϕ21 (right), the SAR endolysins (closed green symbols) accumulate in the IM in an inactive form, anchored by a weakly hydrophobic TMD (green stub) alongside the pinholins (red ovals). See introduction for details on the pathways. b) Different spanin genetic architectures from the phages λ (embedded), P2 (overlapped), T4 (separated) and T1 (unimolecular). The rectangles, drawn to scale and labeled with appropriate names are color-coded to represent different lysis genes in each phage: red (holin), green (endolysin), blue (i-spanin), yellow (o-spanin) and cyan (u-spanin), while the uncolored rectangles represent genes of unrelated or unknown function. The spanin genes are also labeled with their spanin type on top of the gene in each case to highlight the genetic architecture. Note that color does not indicate sequence similarity; in fact, among all the genes depicted, only the K and R endolysin genes share any detectable similarity.

The importance of these genes for lysis had been previously overlooked because in the context of the shaker flask, the fragile spherical cells were destroyed by shearing forces [114]. Thus, in cultures aerated by shaking, spanins are not required for lysis unless the OM is stabilized by addition of millimolar levels of divalent cations to the media. One surprising implication of these findings was that the OM alone could withstand the internal osmotic pressure in the absence of the PG layer, a conclusion also reached in studies of antibiotic-treated *E. coli* cells [171].

Biochemical and genetic studies revealed that Rz to be a class II (N-in, C-out) membrane protein in the IM, with a periplasmic domain dominated by predicted alpha helical domains (Fig. 2.2a,b) [52, 115, 134]. The alpha helices have a high propensity to form coiled-coils resulting in the periplasmic domain of Rz to be essentially divided into two coiled-coil domains, CC1 and CC2, connected by a flexible linker region (Fig. 2.2a). Rz1 was shown to be an OM lipoprotein lacking any detectable secondary structure, presumably due to the 10 Pro residues in its mature 40 amino acids (aa) periplasmic domain (Fig. 2.2a,b). It was demonstrated that Rz and Rz1 form a complex through C-terminal interactions [115, 134]. Since this complex spans the periplasm, the two proteins were named as subunits of a spanin complex, with Rz as the prototype i-spanin (IM subunit) and Rz1 as the prototype o-spanin (OM subunit) [129] (Fig. 2.2a; left). Genes arranged like *RzRz1* were categorized as two-component spanins (2CS) and found to be common in phages of Gram-negative hosts [116], strongly supporting the general role of the spanin complexes in OM disruption and thus, lysis. Furthermore, i-spanin/o-spanin genes could also be found in two other architectures apart from the completely embedded architecture in lambda: overlapped, where the o-spanin gene extends beyond the i-spanin reading frame, and separated, where the genes do not overlap (Fig. 2.1b). In the same study, it was found that in a few phages, including the paradigm phage T1, the last gene in the lysis cassette (Fig. 2.1b) encoded a single

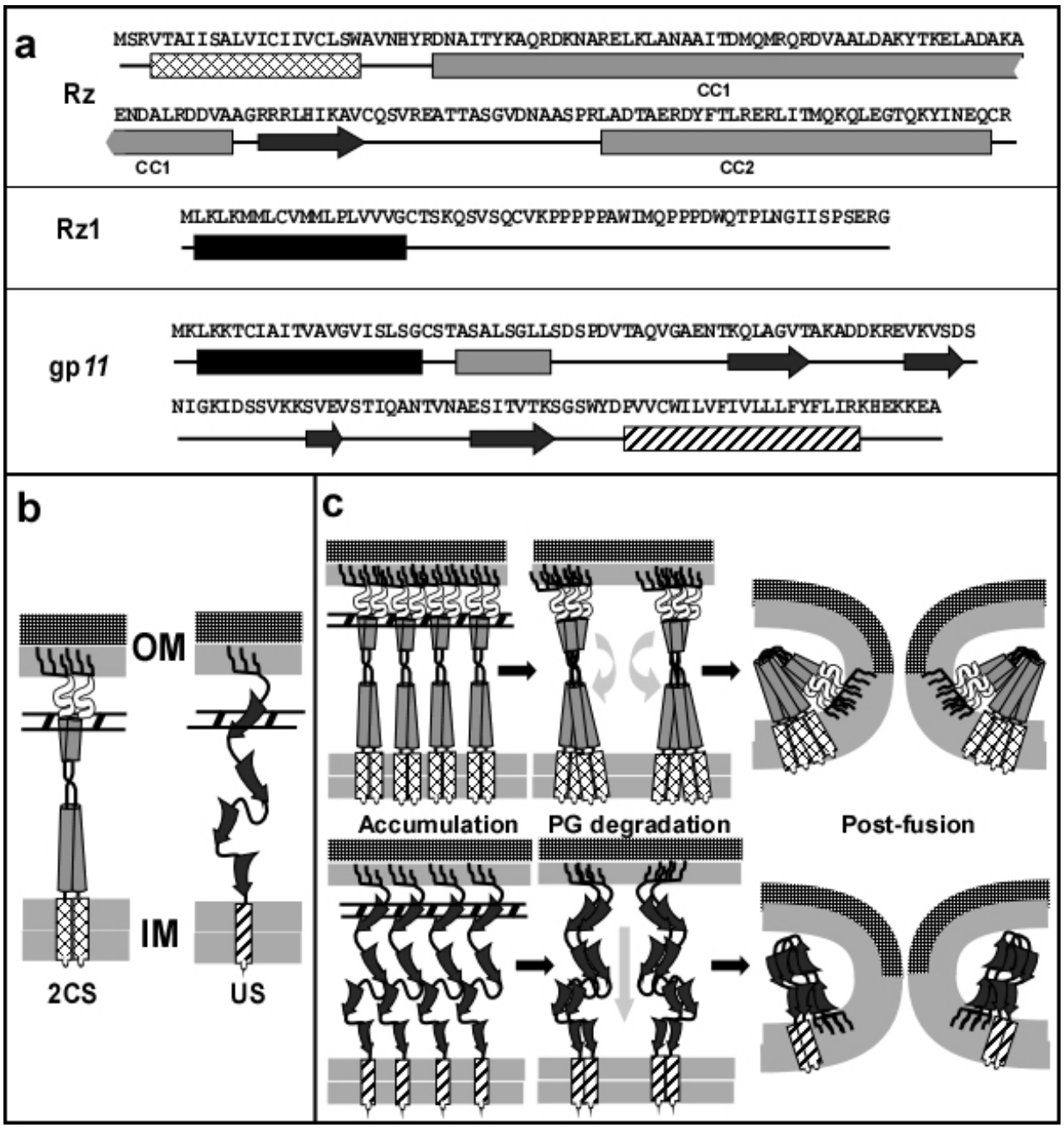
protein with an OM-lipoprotein signal and a C-terminal TMD (Fig. 2.2a; right). The T1 gene, *11* was shown to complement the *RzRz1* lysis defect, indicating that gp*11* also played the same role in lysis; i.e., disruption of the OM. Thus, gp*11* was designated as the prototype unimolecular spanin (u-spanin). Primary structure analysis indicated that the periplasmic domain of gp*11* was predicted to be comprised mainly of beta sheets, in contrast to the coiled-coil alpha helices predicted in i-spanins (Fig. 2.2a, b).

Molecular function of two-component spanins

Recent studies addressing the molecular mechanism by which the two-component spanin systems function have led to a model for spanin function [117] in which the spanin complex effects removal of the OM barrier by fusing it to the IM (Fig. 2.2c); thus, spanins mediate a topological solution, rather than a degradative one, for the last step of lysis. The most compelling results supporting the fusion model came from experiments with *E. coli* spheroplasts [117]. Spheroplasts expressing *Rz* were shown to undergo efficient fusion with spheroplasts expressing *imRz1*, a missense allele of *Rz1* in which the lipoprotein OM-localization signal was altered to cause retention in the IM. In contrast, lysis-defective missense alleles of either *Rz* or *Rz1* promoted adhesion of spheroplasts but did not support fusion, indicating that these mutants are blocked at a step after complex formation but before fusion.

A more recent genetic analysis revealed the functional significance of different domains in Rz and Rz1[134]. Mutations in the coiled-coil domains CC1 and CC2 of Rz rendered the spanin complex non-functional, but the linker domain was shown to act as a flexible spacer, permissive to insertions of Gly-Ser oligopeptide repeats. It was also determined that a proline-

Figure 2.2 Predicted secondary structures, topology and model for function of two-component spanins and unimolecular spanins a) Predicted secondary structures of lambda i-spanin Rz, o-spanin Rz1 and T1 u-spanin gp*11*. The TMDs in Rz and gp11 are shown as crossed and hatched rectangles respectively, while the lipoylation signal sequences in Rz1 and gp11 are shown as black rectangles. Grey rectangles and black arrows indicate predicted alpha helical and beta sheet domains respectively. Coiled-coil domains CC1 and CC2 are connected through a flexible linker region. b) Cartoon representation of the topology of two-component spanin prototype from phage λ and the u-spanin from phage T1. In phage λ, i-spanin Rz is embedded to the IM by an N-terminal TMD (crossed rectangle) and has a periplasmic domain that constitutes two alpha helices (grey cylinders) connected by a linker, predicted to form coiled coils. The o-spanin Rz1 is attached to the inner leaflet of the OM via three fatty acyl chains (black lines) and has a periplasmic domain (white) predicted to be unstructured. The i- and o- spanins interact through their C-termini to form the spanin complex, linking the IM and OM through the PG meshwork. In T1, the u-spanin gp11 is attached to the inner leaflet of the OM, by the three fatty acyl chains (black lines) at the N-terminus and to the inner membrane, through the C-terminal TMD (hatched rectangle). The periplasmic domain of gp11, predicted to be mainly extended beta sheets (black arrows), connects the IM and OM through the PG meshwork. c) Cartoon representation of the model for the function of two-component (top panels) and unimolecular (bottom panels) spanins. The spanin complexes accumulate within the constraints of the PG meshwork, connecting the IM and OM (left panels). Once the PG is removed by the endolysin, these complexes are free to undergo lateral diffusion and conformational changes (middle panels). These events bring both the membranes together, leading to fusion of the IM and OM (right panels) and release of the phage progeny through the fusion pore. The model presented in the top panel has already been presented in a previous report by Rajaure et al. [172].



rich region (PRR) in Rz1 was mutationally sensitive, with most non-functional mutations mapping to the central penta-proline stretch. Additionally, the N-terminus region of mature Rz1 contained a flexible linker region between the lipoylated Cys and the PRR; Gly-Ser repeat additions to this domain did not affect spanin function. Both coiled-coil domains and PRR

domains are known to be a common feature of well-studied membrane fusion motifs in the existing literature [138, 153]. Furthermore, the genetic analysis also provided clues about covariance at different positions and thus potential interaction sites between Rz and Rz1. While characterizing the non-functional spanin mutants, it was observed that complex formation was unaffected for most *Rz* alleles but was defective for Rz_{E150G} [134]. However, both the lysis and spanin complex-formation defects were suppressed by introducing an R59E change in Rz1, suggesting an interaction between the E150 residue in Rz and the R59 residue of Rz1. In addition, exchanging the charges at these positions, i.e. Rz_{E150R} and $Rz1_{R59E}$ also restored complex formation and lytic function, indicating the interaction occurred through a salt bridge between these residues. To follow up, a suppressor analysis was done for a number of lysis-defective spanin mutants in order to detect more contact points within the Rz-Rz1 hetero-tetrameric complex [135]. Surprisingly, most of the suppressors clustered along the juxtamembrane region of CC1 in Rz and were not allele-specific. These mutations, mostly polar insertions into the hydrophobic core of CC1, were proposed to disrupt the stability of the juxtamembrane region, leading to a conformational state that overcomes the fusion block created by the primary mutation.

Since the first study describing the classification of spanins into two-component spanin and u-spanin systems [116], many new phage genomes have been deposited in the databases [19, 173, 174]. Here we present a comprehensive compilation of the spanin genes now identifiable in the public databases and relate features defined by this analysis to the proposed model for spanin function.

Results and discussion

Identification of spanins and implications for automated phage annotation

We restricted our search to RefSeq genomes of phages of Gram-negative hosts. Spanins have only been recently characterized, compared to the long history of studies on other lysis proteins, and essentially all experimental work has been done with the lambda *Rz* and *Rz1* genes only. Thus, it was not surprising to find that less than 20% of these genomes had both an i-spanin and o-spanin identified; most of these were sequence homologs of the two-component spanin systems from paradigm phages lambda Rz/Rz1, P2 LysB/LysC or T4 PseT.3/PseT.2, due to the over-representation of lambdoid, P2 and T4-like phage genomes in the database. The requirement for an outer membrane lipoprotein signal for both two-component spanins and u-spanin systems suggested that an automated strategy based on identifying genes with such signals could be implemented. However, our initial attempts along this line were frustrated because in most cases, the o-spanin genes in the embedded and overlapped architectures were not annotated as CDSs, reflecting a strong bias against overlapped genes in the most commonly used gene-calling programs like, Glimmer [175, 176] or GeneMark [177-179]. The second problem was that many CDSs in phage genomes have misidentified start codons, again probably linked to the bias against overlapped genes in gene-calling programs and the tendency of phage genes to overlap [180]. Since signals necessary to localize proteins to any location in the envelope, including the IM, periplasm, and OM, are always at the N-terminus of a protein, misidentified start codons are extremely problematic for the identification of spanins and, indeed, all lysis proteins. Consequently, we implemented a manual search protocol, shown in Fig. S2.1, augmented by specialized work-flows constructed at the Center for Phage Technology Galaxy

instance [181]. Using this protocol, we interrogated 677 genomes of dsDNA phages of Gram-negative hosts and found 528 two-component spanins and 58 u-spanins, as described in Table 2.1a. The rest of the 91 genomes did not possess any potential spanin candidate genes that met our eligibility requirements for the membrane localization signals. Of the 528 two-component spanin systems, 182, 228 and 118 belonged to the embedded, overlapped and separated architectures, respectively. Table S2.1 contains all the coordinates, sequences and other features of the spanins identified in this survey. (Table S2.1 serves as the initial basis for a continuously updated SpaninDataBase (SpaninDB) at the Center for Phage Technology website [182]). The results justified our decision for manual annotation. Of the 528 genomes with two-component spanin systems, the CDS for the o-spanin had errors that would preclude automated annotation in 260 cases (196 entirely missing CDSs and 64 CDSs with incorrect start sites) (Table 2.1b and S2.1). The problem was less severe with the i-spanins (13 missing, 34 with the incorrect start site) and u-spanins (4 with incorrect start site). These findings suggest that phage-specific algorithms for gene calling are needed before accurate automated analysis of phage genomes can be practical. Efforts to modify the concerned RefSeq accessions of CDS with such errors are in progress, in collaboration with NCBI.

As a primary analysis, we first inspected the overall length distribution of spanin complexes, and their secondary structure distribution as predicted by Jpred (Table S2.2). It can be expected that since the spanin complex formed by the interaction of the periplasmic domains needs to span the entire periplasm, there would be length restrictions for the number of residues that would be required to physically connect the membranes. However, the potential to adopt various secondary structures and the dependence of periplasmic width on host and environmental conditions, would allow for a varied range of lengths. This was very evident from the

periplasmic length profiles of 2CS, both the periplasmic domains put together, ranging from as short as 62 residues for the spanins from phiP27 to 300 residues for the spanins from Marshall (Table S2.2). The differences were also noticeable in secondary structure profiles; the first major observation from these predictions was that, in contrast to the unstructured character of the periplasmic domain of Rz1, several embedded spanins were predicted to have a significant structural component. For example, the HK97 o-spanin showed as high as 47% alpha helical character, while the o-spanin from phage HK225 showed as high as 28% beta sheet character. The beta sheet character was more prominently found in the separated o-spanins compared to the overlapped and the embedded architectures. Given the proposed role of the coiled-coil domains in lambda spanin function, we asked if the coiled-coil domains, as predicted by the tool Pepcoil, were conserved among other 2CS (Table. S2.3). As expected from the secondary structure predictions, the variation of coiled coil character was also very high, with more than 120 of the i-spanins having no coiled-coil predictions at all, and more than 250 i-spanins having one coiled-coil domain. All these differences in structural domain organization of the spanins hint that the eventual goal of spanin function, i.e. membrane fusion of the OM with IM, may be achieved in different ways by different spanin complexes, not necessarily only through coiled-coil domain interactions as proposed in lambda. In conjunction with this theory, we did not observe conserved periplasmic lengths or secondary structure distributions for different phages infecting the same host (Table. S2.2). Given that the periplasmic width changes depending on the type of the host and growth conditions [50, 183], it cannot be ruled out that spanin complexes of different length might be sequestered to different parts of the periplasm. Extensive structural studies will need to be done to understand the detailed mechanistic differences between spanins

of different structures and if any conformational changes occur between free and complex bound spanins.

We also examined the register and position of the o-spanin gene with respect to the i-spanin gene, especially for the embedded and overlapped architectures (Table S2.4). All the previously reported o-spanins were found in the +1 reading frame of their respective i-spanins [116], suggesting that the codon mix available from +1 frameshifts was required for maintenance of o-spanin function. However, in this survey, we found that the o-spanin occupied the -1 frame of the i-spanin in 116 cases (Table S2.4). Considering the over-representation of homologs of the lambda, T7 and P2 spanins, which are the experimentally tested embedded (lambda and T7) and overlapped (P2) two-component spanin systems and all of are +1 architectures, we conclude that there is no significant bias to either the +1 or -1 reading frame for o-spanin evolution. Analysis of the relative position of the o-spanin with respect to the i-spanin genes showed that in ~70% of the embedded spanins, the o-spanin started at ~0.4-0.5L and ended at 0.9-1L, where L is the length of the corresponding i-spanin (Table S2.3). Combined with our findings from lambda genetics [134], it can be interpreted that the positioning of the embedded o-spanin gene would need to be conserved, as the shared DNA would need to accommodate for the homology of both the CC2 domain of the i-spanin and the C-terminal region of the o-spanin. Since the extreme C-terminus does not seem to be involved in interaction sites for spanin complex formation, the last 0.1L of the i-spanin is subject to variability.

Diversity of the two-component spanin systems

To assess the sequence diversity of spanins, we grouped the spanins into families using BLASTCLUST on the CPT Galaxy platform [184], defining families such that every member

shared ≥ 40% identity over ≥ 40% of the length of the periplasmic sequence with every other member. Focusing on the periplasmic domain avoids the low complexity regions within the IM and OM lipoprotein localization signals, which are necessary for membrane anchoring only and have no functional significance [115, 134]. This approach resulted in the 528 two-component spanins systems being grouped into 157 i-spanin (99 singletons) and 136 o-spanin (65 singletons) families and the 58 u-spanins, into 13 (6 singletons) families. BLASTCLUST uses BLOSUM62, a "deep" scoring matrix that requires long sequence alignments [185, 186]. Thus, all the sequences with a periplasmic domain shorter than 50 aa were combined with the singletons from the BLASTCLUST analysis and manually clustered into new or existing families, as per our definition. This eventually resulted in 143 i-spanin (80 singletons) families, 125 (54 singletons) o-spanin families and 13 u-spanin (6 singletons) families (Table 2.2, 2.3 and 2.4). A substantial component of the 2CS collection was biased towards homologs of the spanins of the lambdoid phages and T4-like phages. The largest i-spanin family, represented by lambda Rz had 83 members, and the largest o-spanin family represented by T4 PseT.3 had 47 members while a majority of the u-spanins fell under the T1 family with 28 members.

The difference between the number of i-spanin and o-spanin families (143 versus 125, respectively) was intriguing. This could be partly attributed to the difference in lengths and aa composition of the spanin counterparts, considering the way BLOSUM62 similarity scores are calculated [186]. Long sequence alignments are easier to achieve in the i-spanin components than the o-spanin components; the i-spanin component is longer than the corresponding o-spanin component in 489 of the 528 two-component spanins described in this study. The difference in number and composition of i-spanin and o-spanin families also reflects the distinct evolutionary paths of spanin components, independent of each other. For example, in the case of lambda

spanins, the Rz family had 83 members, whereas the respective o-spanin counterparts were distributed into 7 different families, most of them spread across three families HK97 (32 members), lambda (28 members) and HK620 (14 members). This essentially means that the 83 i-spanins could be arranged into alignments with statistically significant scores, enough to cluster them into a family according to our criteria, whereas the alignments of their o-spanin counterparts did not lead to scores sufficient for all of them to be included into one single family. Instead, the alignments of the 83 o-spanins could be divided into different groups, where the scores within each of the individual group allowed them to be clustered into a family. This suggested that the members within these o-spanin families were more closely related to each other than members from other families, in terms of evolutionary distance. This observation was surprising, as it indicated that these o-spanin sequences diverged more than i-spanin sequences, even though they were totally embedded within the latter. This means that the segment of the i-spanin gene that contains the coding sequence for the periplasmic domain of the o-spanin in the +1 frame would have to be more permissive for changes, while the rest of the gene was relatively conserved. This interpretation was consistent with the findings from our recent genetic analysis of the lambda spanins [134]. The DNA encoding the periplasmic domain of Rz1 is shared with the genetically flexible linker region in Rz. Notably, even though the genetic study was done on a synthetic separated pair of Rz-Rz1, the mutational clusters of one gene corresponded to the mutationally silent region in the other, highlighting the differential evolutionary pressures on the embedded spanins.

Interestingly, when we repeated the clustering with 100% sequence identity over 100% sequence length (full length, including the signal sequences), we found that there were 46 sets of 2CS, spread across the three different genetic architectures, and 3 sets of u-spanins accounting

for 113 and 6 sequences respectively, each set having entirely identical sequences (Table 2.5). Remarkably, a few of the identical spanins came from phages that infected different hosts, which reinforces the mosaic nature of phages [20]. For example, Enterobacteria phages T3 and T7M share identical embedded spanins with Yersinia phages R, YpP-Y and YpP-R. The unexpectedly high frequency of identical spanins was probably a byproduct of our strategy of using a BLAST-DB of existing spanins to hunt for new spanins (Fig. S2.1). Nonetheless, this strategy helped us identify spanins in new phage genomes, even if identical to already existing spanins, where they had not been identified before or even annotated. Even with the presence of a high number of identical spanins, we still observed 80 i-spanin and 54 o-spanin singleton classes, or ~ 55% and 43% of the total number of families of i-spanins and o-spanins respectively having only one member, highlighting the extreme diversity of this functional class of lysis proteins.

Conserved functional domains and covariance

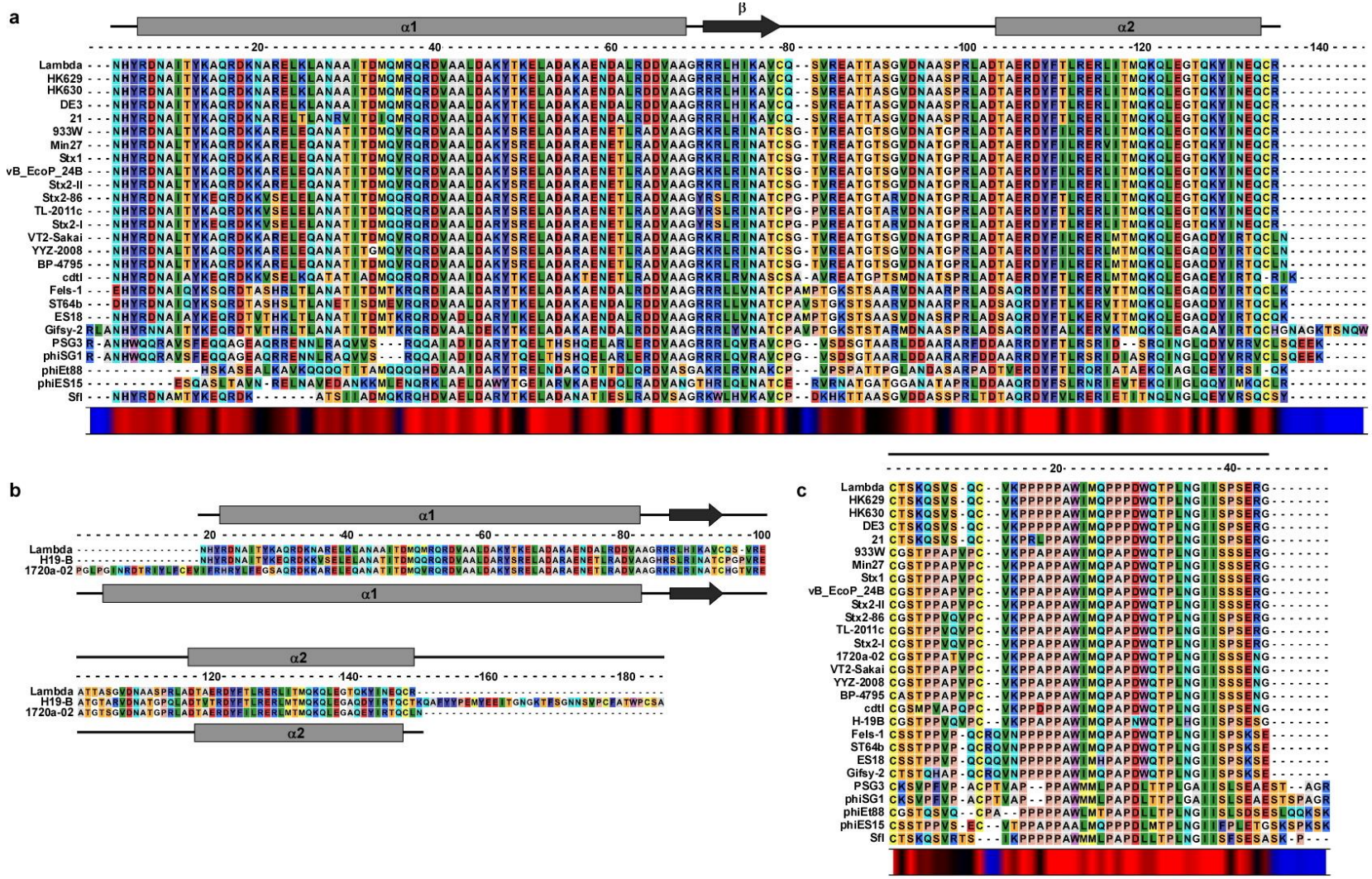
As a follow up on the genetic analysis of the lambda spanins [134], we investigated the alignments of the different spanin family members to see if the mutationally sensitive regions and the functional domain organization were conserved. We hoped to visualize covariance amongst residues that could potentially be involved in primary and secondary site interactions, important for different stages of the spanin operation. We performed multiple sequence alignments of the i-spanin and o-spanin sequences of all families, based on both i-spanin and o-spanin clustering. However, in this report we limit ourselves to discuss the alignments of the one family from each genetic architecture that gave us the most significant insights into conserved regions.

For the embedded architecture, we analyzed the alignments and secondary structure organization of sequences from 26 members of the lambda o-spanin family (Table 2.3) and their i-spanin counterparts. The alignments revealed highly conserved regions in both spanin components that were consistent with the mutationally sensitive positions identified in the experimental genetic analysis [134]. The regions of alignment of i-spanins corresponding to mutationally sensitive regions of α1 (between residues R59 and A88) and α2 of lambda Rz (between residues D126 and Y147), were conserved while the linker region (between residues Q100 and S115) showed variability in composition, in accordance with the proposed flexibility from the genetic studies (Fig. 2.3a). The periplasmic domains from i-spanins of 1720a-02 and H19-B were remarkably longer than the other members. Upon alignment, we noticed that this was a result of addition of long aa stretches at the extreme N-terminus for 1720a-02 (19 residues) and C-terminus for H19-B (37 residues) (Fig. 2.3b). Secondary structure predictions showed that the N-terminal additional segment in 1720a-02 is predominantly alpha-helical albeit with no coiled-coil character, whereas the C-terminal additional segment in H19-B is mostly unstructured. We hypothesize that the segments are probably dispensable and not involved in interactions necessary for spanin function. Based on our findings with lambda spanins [134], we proposed the N-terminal region of 1720a-02 might act as a long helical linker connecting the α1 to the IM and the C-terminal domain in H19-B might act as flexible loop allowing the distal end of α2 to interact with the o-spanin. Upon inspecting the o-spanin alignments (Fig. 2.3b), we found the penta-proline stretch in lambda Rz1 was conserved in all the other o-spanin members, except for P34, which was also the only position in that stretch where no lysis-defective mutations were found [134], underscoring the permissive nature of the position. In addition, the linker region between the N-terminal lipoylation Cys and the penta-proline stretch was relatively

variable, gaining more prolines which would add to the unstructured character of this region. This agreed with our genetic analysis [134] which showed that the linker region was flexible and allowed for the addition of Gly-Ser repeats. The o-spanin sequences from PSG3, phiSG1, phiEt88, phiES15 and SfI all had longer, and variable C-terminal ends compared to the other members. However, these additional segments might not be relevant to spanin function as the lambda genetics showed the extreme C-terminus, though involved in interactions with the i-spanin counterpart, is dispensable [134].

The lambda spanin family alignments also revealed that the Cys residues at position 29 in the o-spanin and 99, 152 in the i-spanin are unchanged in almost all members, except for cdtI, phiEt88, and SfI. cdtI and phiEt88 lack the C152 in the i-spanin, whereas SfI lacks the C29 in the o-spanin. However, we expect these spanins to still be functional as the cysteine substitutional analysis of lambda Rz-Rz1 showed that the presence of just one Cys in either of these positions was necessary and sufficient for spanin function [129]. It was notable that the C99 is a constant feature in all these i-spanins despite being dispensable for spanin function, indicating there might be some advantage in terms of structural stability conferred by the intermolecular disulfide bond at that position. Furthermore, the alignments also supplemented the findings of our suppressor analysis [135]. For examples, secondary site suppressors Rz_{Y33C} and Rz_{A47S} were isolated against the lysis-defective Rz_{V61A} allele, while Rz_{D38G} was isolated against Rz_{L64H}. We examined these specific positions in the alignments and found that the i-spanins from both PSG3 and phiSG1, which contained an Ile instead of Val at position 61 and Ile instead of Leu at position 64, also had substitutions at positions 33 (Y33→F), 38 (D38→G) and 47 (A47→L). This strongly suggested covariance at these positions, but it is not clear if these sites are involved in interaction, as there seems to be no allele specificity. The covariance between $Rz1_{R59}$ and Rz_{E150}

Figure 2.3 Sequence alignments of lambda spanin families. (a) and (c) show alignments of i-spanin and o-spanin sequences (periplasmic domains only) of the lambda family respectively. The alignments of i-spanin sequences from 1720a-02 and H-19B with respect to the lambda i-spanin are shown separately in (b) as their i-spanin sequences were significantly longer compared to other members. Secondary structure predictions for respective spanins from lambda are shown above each alignment. The secondary structure predictions for i-spanins from H19B and 1720a-02 are shown above and below respectively, in the additional alignment in 3b. Grey rectangles indicate alpha helices and black arrows indicate beta sheets. Conservation at different positions is shown below with colors (Red indicates maximum conservation and blue indicates least conservation). A scale indicating the relative position of the residues and the approximate size of various domains is shown above the alignments.


a
α1
β
α2
Lambda
HK629
HK630
DE3
21
933W
Min27
Stx1
vB_EcoP_24B
Stx2-II
Stx2-86
TL-2011c
Stx2-I
VT2-Sakai
YYZ-2008
BP-4795
cdtI
Fels-1
ST64b
ES18
Gifsy-2
PSG3
phiSG1
phiEt88
phiES15
Sfl
b
α1
α2
Lambda
H19-B
1720a-02
c
Lambda
HK629
HK630
DE3
21
933W
Min27
Stx1
vB_EcoP_24B
Stx2-II
Stx2-86
TL-2011c
Stx2-I
1720a-02
VT2-Sakai
YYZ-2008
BP-4795
cdtI
H-19B
Fels-1
ST64b
ES18
Gifsy-2
PSG3
phiSG1
phiEt88
phiES15
Sfl

was also quite evident, adding to our salt bridge interaction findings. Using the alignments, we could identify other covariance patterns, such as $Rz1_{W46\rightarrow L}$- $Rz_{E142\rightarrow N}$ and $Rz1_{V26\rightarrow A}$ -$Rz1_{P56\rightarrow S}$. However, it remains to be tested if these residues are involved in interactions relevant for spanin function.

For the separated architecture, we analyzed the alignments of sequences from 40 members of the T4 i-spanin family and their o-spanin counterparts, shown in Fig. 2.4a,b respectively. Secondary structure predictions for the T4 spanins are significantly different from the lambda spanins. The periplasmic domain of T4 i-spanin, PseT.3 essentially consists of two alpha helices, designated as α1 and α2 (Fig. 2.4a), connected by a very short linker region. However, unlike lambda Rz, only α1 has the propensity to form coiled-coils. In comparison to the unstructured lambda Rz1, the T4 o-spanin PseT.2 contains a β1−α−β2 domain organization, a flexible linker connecting the OM to the central beta sheet domain, followed by an alpha helix motif, and a very short beta sheet, while the extreme C-terminus has no predicted structural content. The alignments showed significantly conserved regions as well secondary structure organization in both spanin components. Like lambda, the T4 o-spanin family seemed to have a highly variable linker region between the N-terminal Cys linking to the OM and the central beta sheet domain (Fig. 2.4b). The β1−α−β2 stretch ranging from W47 to R90 was greatly conserved. Unlike the lambda o-spanin family, the T4 o-spanin family lacks any proline rich region or proline stretches. Still, the Pro residues at positions 21, 31, 34 and 61 were conserved in almost all the members, suggesting these residues might be involved in the fusion activity. Moreover, both the Cys residues at positions 87 and 98, were unchanged in all members, emphasizing the role of the intra- and intermolecular disulfide bonds at these positions for spanin function. On the other hand, the i-

spanin alignments showed a conserved α1 and α2 domain organization, even though the N-terminus of α1 (connecting to the IM) region was extremely variable in composition. Even with the extreme variation in α1, α2 was conserved across all the members, starting around L96 up to the C-terminal end, implying this domain could be involved in potential interaction with the conserved regions of the o-spanins. Upon inspection of these alignments for intra- and intergenic covariance, we observed that the a negatively charged residue at position 92 in the o-spanin was almost always associated with the presence of a K117 in the i-spanin, suggesting these two residues could be involved in polar interactions, probably through a salt bridge, at some stage of the spanin pathway.

For the overlapped architecture, we analyzed the alignments of sequences from 24 members of the Jersey i-spanin family (Table 2.2; we omitted the sequences from FSL_SP-031, -038, and -049 as they belonged to the separated genetic architecture. Alignments including these sequences are discussed in the next section.) and their o-spanin counterparts as shown in Fig. 2.5a,b. The alignments showed fairly conserved sequences and secondary structure organization spread across the periplasmic domains of both the spanins. The Jersey i-spanin family could be broadly classified into two subfamilies, based on the length of the periplasmic domains and the predicted secondary structure profiles. The first subfamily (Jersey to SE2 in Fig. 2.5a) had a secondary structure organization of two alpha helices connected by a linker region. But unlike the unstructured linker as seen in lambda, the linker region of these i-spanins had 2 predicted beta sheets, the first instance of i-spanins showing significant beta sheet secondary structure character. The second subfamily (EP23 to SO-1 in Fig. 2.5a), shorter in length, had a similar two alpha helix domain organization, with only one beta sheet motif predicted in the linker region.

Figure 2.4 Sequence alignments of T4 spanin families. (a) and (b) show alignments of i-spanin and o-spanin sequences (periplasmic domains only) of the T4 family respectively. Secondary structure predictions for respective spanins from T4 are shown above each alignment. Grey rectangles indicate alpha helices and black arrows indicate beta sheets. Conservation at different positions is shown below with colors (Red indicates maximum conservation and blue indicates least conservation). A scale indicating the relative position of the residues and the approximate size of various domains is shown above the alignments.

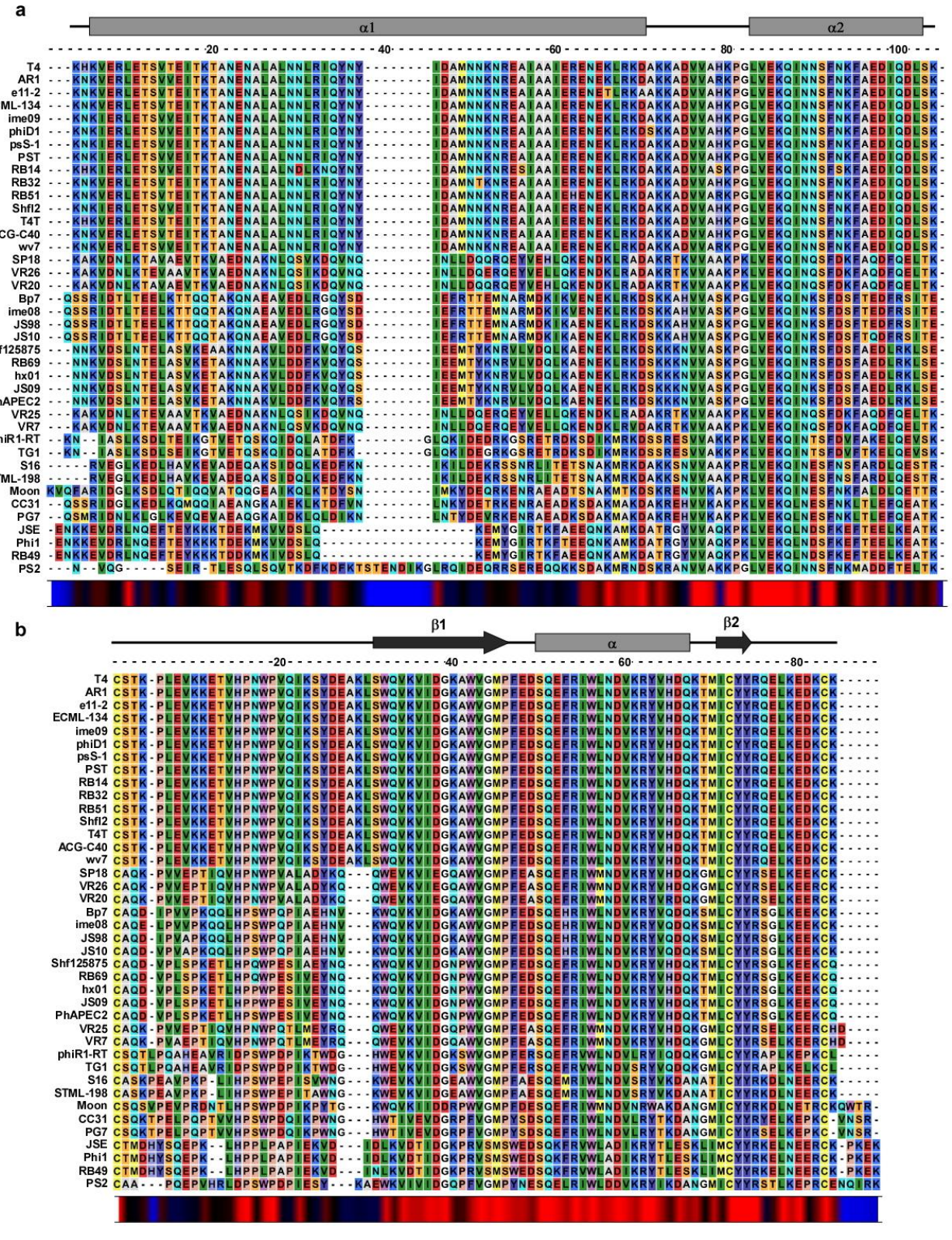
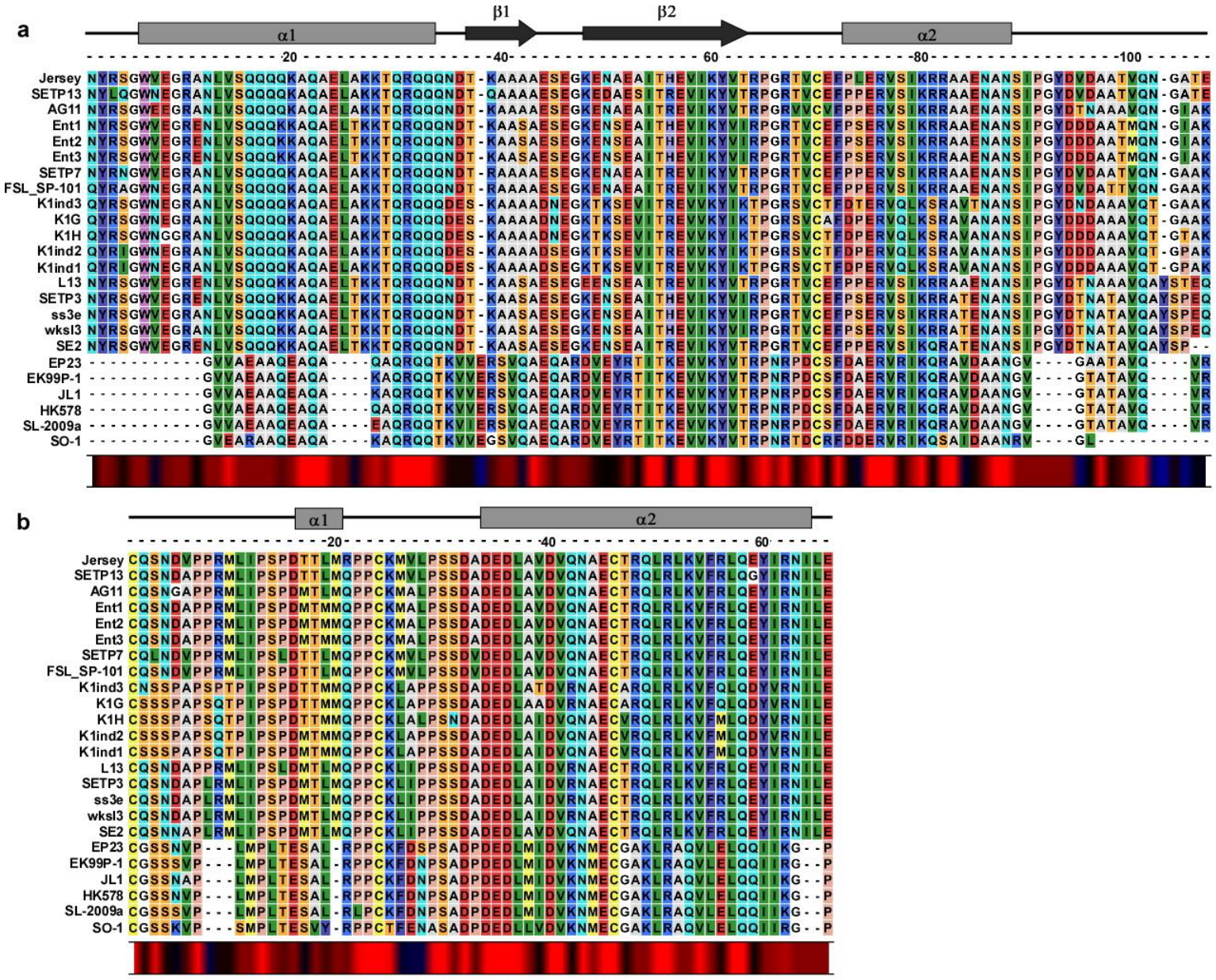

a
α1
α2
b
β1
α
β2

Figure 2.5 Sequence alignments of Jersey spanin families. (a) and (b) show alignments of i-spanin and o-spanin sequences (periplasmic domains only) of the Jersey family respectively. Secondary structure predictions for respective spanins from Jersey are shown above each alignment. Grey rectangles indicate alpha helices and black arrows indicate beta sheets. Conservation at different positions is shown below with colors (Red indicates maximum conservation and blue indicates least conservation). A scale indicating the relative position of the residues and the approximate size of various domains is shown above the alignments.


a
α1
β1
β2
α2
20
40
60
80
100
Jersey
SETP13
AG11
Ent1
Ent2
Ent3
SETP7
FSL_SP-101
K1ind3
K1G
K1H
K1ind2
K1ind1
L13
SETP3
ss3e
wksl3
SE2
EP23
EK99P-1
JL1
HK578
SL-2009a
SO-1
b
α1
α2
20
40
60
Jersey
SETP13
AG11
Ent1
Ent2
Ent3
SETP7
FSL_SP-101
K1ind3
K1G
K1H
K1ind2
K1ind1
L13
SETP3
ss3e
wksl3
SE2
EP23
EK99P-1
JL1
HK578
SL-2009a
SO-1

The extreme C-terminal end of both the subfamilies was highly variable as well as unstructured, signifying the α2 region, not the extreme C-terminus, was probably involved in interaction with the i-spanin. Interestingly, the first subfamily also contained two polyQ stretches. PolyQ repeats are known to facilitate and stabilize coiled-coil interactions [187], and could potentially affect spanin function in this case by strengthening homotypic interactions between adjacent spanin complexes. All the o-spanins had a conserved 2 alpha helical prediction (denoted as α1 and α2 in Fig. 2.5b) connected by a flexible linker, in addition to the flexible linker region connecting the N-terminal Cys to the central short helix. The position and composition of α1 varied in different members, while α2 was relatively more conserved, again suggesting involvement in interactions with i-spanin. Both Cys at positions 45 and 67 remained constant, hinting at the potential role of intermolecular disulfide bonds in spanin function here. Prolines were also conserved in different positions of both the unstructured linker regions.

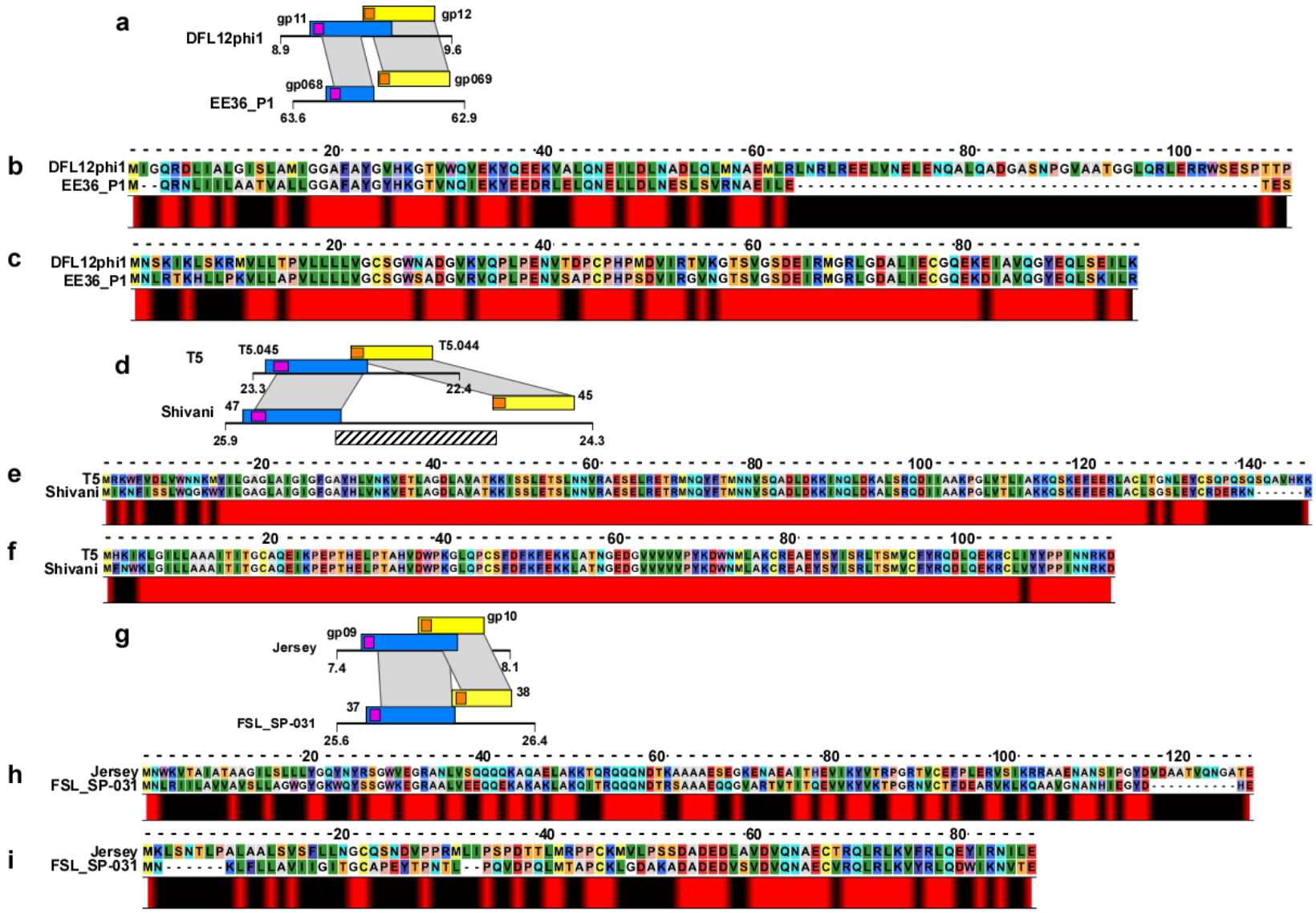
Evolution of the two-component spanin genetic architecture: caught in the act

In a recent study of phages isolated against *Xylella* and *Xanthomonas*, we described a new phage type, represented by the phage Nazgul [188], consisting of 7 phages; Nazgul, Redjac, Enc34, Chi and AH2, Sano and Salvo. While characterizing these phages, we identified and grouped the spanin sequences into 2 families, one consisting of the phages Sano and Salvo, and the others including phages Nazgul, Redjac, Enc34, Chi and AH2. Interestingly, while Salvo had its two-component spanins in an overlapped architecture, the spanin gene pair in Sano was separated. Similarly, the spanins of AH2 were separated, while the respective homologs from other members of the family, Nazgul, Redjac, Enc34 and Chi, were all overlapped. Not only was this the first report showing spanin gene families with members across different architectures,

but also the alignments of these spanin sequences from different architectures hinted at the evolution process of the two-component spanin architecture. While almost all the two-component spanin families were restricted to a single architectural class based on our grouping strategy, we found three i-spanin families and seven o-spanin families that had members from both overlapped and separated architectures. For further analysis, we chose the three families represented by DFL12phi1, T5, and Jersey, where both the i- and o-spanin families had members from the overlapped and separated architectures and hoped to predict the potential evolution pathways that could transition from one genetic architecture to the other.

The first family consists of the overlapped spanins from DFL12phi1 and the separated spanins from EE36P1. The spanins in DFL12phi share 116 bp, whereas the spanins in EE36P1 are separated by 17 bp (Fig. 2.6a). Upon comparing the alignments and secondary structure predictions (Fig. 2.6b,c), we observed that the i-spanin from EE36P1 has a conserved N-terminal alpha helix $\alpha 1$ domain like the DFL12phi i-spanin, but the entire C-terminal stretch including the second predicted alpha helix $\alpha 2$ is missing. The o-spanins are conserved all throughout the periplasmic domain, maintaining the secondary structure domains as well. The T5 family, consisted of nine members, of which seven were of the overlapped genetic architecture; EPS7, Stitch, SPC35, bV_EcoS_AKFV33, vB_EcoS_FFH1, DT57C and T5 and the other two, phiR201 and Shivani, were of the separated genetic architecture. To simplify the analysis, we chose T5 to represent the overlapped class and aligned its two-component spanins with the sequences from Shivani. Upon examination, we found that in Shivani and phiR201 the spanin pair are separated by ~700 bp that includes a gene encoding for a putative GIY_YIG homing

Figure 2.6 Sequence alignments of spanins sharing homology across different architectures. a, d, g are pictorial representations of the homologous regions of DLF12 and EE36P1, T5 and Shivani. and Jersey and FSL_SP-031 respectively. (b,c) show sequence alignments of periplasmic domains of i-spanin and o-spanin of DLF12 and EE36P1 respectively, (e,f) show sequence alignments of periplasmic domains of i-spanin and o-spanin of T5 and Shivani respectively and (h,i) show sequence alignments of periplasmic domains of i-spanin and o-spanin of Jersey and FSL_SP-031 respectively. We chose to include the full-length sequences in the alignments as the shared DNA region codes for the signal sequence of the o-spanin. Blue and yellow rectangles represent i-spanin and o-spanin respectively, whereas the pink and orange rectangles represent the approximate position of the N-terminal TMD sequence in the i-spanin and the N-terminal lipoylation signal sequence in the o-spanin respectively. Grey shaded region indicates the region of sequence similarity between homologs. Gene names are in bold, all genes are drawn to scale, and their position in the genome in terms of kilobases is denoted by the numbers below the line. Conservation at different positions is shown below with colors (Red indicates maximum conservation and blue indicates least conservation). A scale indicating the relative position of the residues and the approximate size of various domains is shown above the alignments.


a
DFL12phi1
gp11
gp12
8.9
9.6
gp068
gp069
EE36_P1
63.6
62.9
b
DFL12phi1
EE36_P1
c
DFL12phi1
EE36_P1
d
T5
T5.045
T5.044
23.3
22.4
45
Shivani
47
25.9
24.3
e
T5
Shivani
f
T5
Shivani
g
Jersey
gp09
gp10
7.4
8.1
38
37
FSL_SP-031
25.6
26.4
h
Jersey
FSL_SP-031
i
Jersey
FSL_SP-031

endonuclease, whereas the T5 spanins overlap by 70 bp (Fig. 2.6d). The i-spanins are almost identical across the periplasmic domain, except for the extreme C-terminal end, which extends into the o-spanin gene (Fig. 2.6e). The o-spanins are almost identical across the entire length, except for the few residues of the lipoylation signal sequence (Fig. 2.6f). Naturally, the spanins from both the phages have very similar secondary structure predictions. Lastly, the Jersey family consisted of 28 members, of which 25 fell under the overlapped gene arrangement and the 3 are in the separated class (Table 2.2, 2.3). To simplify our alignments, we chose Jersey and FSL_SP-031 to represent the overlapped and separated classes, respectively. The spanins in Jersey overlap by 155 bp, while the spanin pair in FSL_SP-031 have only 8 bp of overlap and are thus separated according to our criteria (Fig. 2.6g). The i-spanins share homology across the entire length, except for the N-terminal TMD and the extreme C-terminal end (Fig. 2.6h). The o-spanins share little to no homology at the N-terminus, including the lipoylation signal sequences, but are highly similar from the central to C-terminal end (Fig. 2.6i).

Each of these three cases of families sharing members from different architecture seems to be hinting at a different mode of evolution between architectures. Comparing the reading frames in the shared region of DFL12phi1 to the corresponding region in EE36P1, it can be hypothesized that an in-del mutation or a premature stop codon at the C-terminal of i-spanin resulted in the separated architecture in EE36P1. However, it remains to be tested if the short i-spanin of EE36P1 is still functional without the extended C-terminus, as in DFL12phi1. In the case of T5 and Shivani, transition may have happened from an overlapped to separated architecture or vice versa depending on if the homing endonuclease gene was inserted or deleted at the C-terminal end of the i-spanin. In either case, the insertion or deletion would need to

happen in a manner such that the newly generated N-terminus of the o-spanin gene still encoded the lipoylation signal sequence. The most parsimonious interpretation of the findings from Jersey and FSL_SP-031 alignments is that the spanin genes started with a separated spanin architecture, and the transition from separated to overlapped architecture involved a duplication of the entire spanin locus, followed by deletion event. The fact that average length of both the spanin genes put together is not more than 250 residues, i.e. ~0.75kb, implies that the duplication is very probable. That would allow the phage to manipulate one copy of the spanin pair, through deletions and other mutations, until the duplicated copy is not only functional but also shorter. Once both the copies are functional, the longer version can be deleted. The variability in the extreme C-terminus of the i-spanin allows for the transition where the o-spanin is integrated in the +1 reading frame. Our suppressor analysis [135] using a synthetic separated pair of Rz-Rz1 illustrated that the embedded architecture did not confer any advantage in terms of functional efficiency. Thus, it seems that the only advantage gained is that of increased coding capacity.

In all the three cases, the o-spanins retain their sequence homology throughout the predicted periplasmic domains of the separated and overlapped proteins, but the i-spanins diverged in the C-terminal region encoded by the DNA containing the actual overlap. This suggests that the interaction between the i-spanin and o-spanin in both the overlapped and separated architectures does not depend on the extreme C terminus of the o-spanin, as was demonstrated for the lambda Rz-Rz1 proteins of the embedded architecture. There were no families that had members from the embedded architecture combined with either overlapped or separated architecture. Nevertheless, a similar argument can be put forth that a duplication followed by deletion process could have generated the embedded gene architecture from the separated arrangement. A more complex pathway for arriving at the embedded architecture could

be envisioned where a mutation in the stop codon of the i-spanin gene allows it to extend until a new stop beyond the end of the downstream o-spanin gene. This would need to be followed by internal deletions within the non-overlapping region in the i-spanin to restore the overall length. The intermediate i-spanin in this pathway could easily be fully functional since the addition of a few random residues at the C-terminus of the i-spanin would not necessarily affect the ability to interact with the o-spanin. Moreover, the unusual amino acid content of o-spanins (e.g., lambda Rz1 has 10 Pro residues in its mature 40 aa length) and lack of required secondary structure could facilitate the generation of new i-spanin segments in the +1 or -1 reading frame relative to the o-spanin. In support of this notion, Rancurel *et al.* [189] have shown that genes arising as out of frame sequences embedded within a gene encoding a structured protein, tend to be enriched in disorder-promoting residues (residues that tend to stay unstructured; e.g. Pro, Glu, Ser, Lys [190]). Considering our findings here, the most parsimonious interpretation would be that the embedded and overlapped architectures evolved in parallel, unrelated pathways from primordial separated genes.

Intermolecular disulfide linkages as a universal feature of two-component spanin systems

In the context of spanins, the work of Berry et al. [129] on the lambda Rz and Rz1 proteins showed that the mature complex had 3 homotypic intermolecular disulfide bonds, mediated by the two cysteines at position 99 and 152 in Rz and the sole cysteine at position 29 in Rz1. Genetic analysis revealed that lytic function was preserved as long as at least one of the intermolecular bonds at Rz C152 or Rz1 C29 was retained. The formation of these intermolecular disulfide bonds was dependent on the host Dsb system, and a model for this dependence has been described [129]. We wanted to determine if periplasmic intermolecular

disulfide bonds could be a common requirement for all two-component spanin function and thus inspected the spanin sequences for cysteines. Indeed, we found that, 507 of the 528 two-component spanins we had identified had at least one cysteine in either the i-spanin or the o-spanin component (Table S2.5a). The maximum number of total periplasmic cysteines for two-component spanins was 8, with a maximum of 6 for the i-spanin and 4 for the o-spanin (Table 2.6). The variation in frequency of cysteines in o-spanins with respect to the number of cysteines in i-spanins is shown in (Fig. 2.7a). The most common combination seemed to be 0 cysteines in i-spanin and 2 cysteines in o-spanin, like PseT.3 and PseT.2 from phage T4. As evident from the correlation graph, as the number of cysteines in the i-spanin increases from 0 to 2, the need for an o-spanin to possess a cysteine or more decreases.

The result from Berry *et al.* [129] that either C29 in Rz1 or C152 in Rz, but not C99 in Rz, was required for the function of the lambda spanin complex indicated that the position of the cysteine, and thus, the covalent disulfide linkage, was important. We hypothesized that the cysteine either needs to be proximal to the interaction site of the C-terminal ends of the spanin components or proximal to the OM attachment site of the spanin complex. We extracted the entire periplasmic sequences of the spanin complexes and analyzed the position of the cysteine closest to the IM and OM attachment sites (Table S2.5b and Fig. 2.7b). The probability of finding a cysteine close to the OM was higher compared to finding one close to the IM. Then we inspected individual spanin components and analyzed the position of the cysteine closest to the C-terminal interaction site and membrane attachment site (IM for the i-spanin, OM for the o-spanin) (Table S2.5c,d and Fig. 2.7c). The cysteines in both the i-spanin and o-spanin components were closer to the C-terminal interaction site, mostly within 40% of the sequence

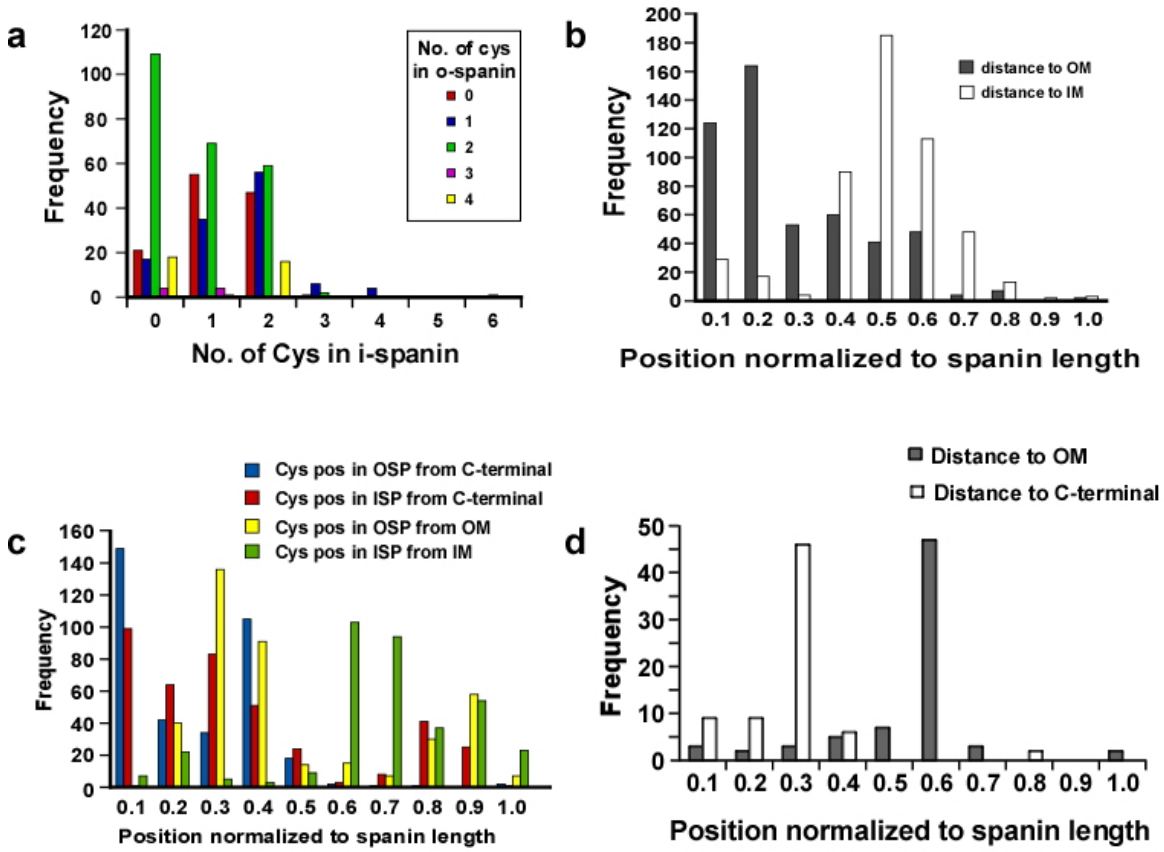


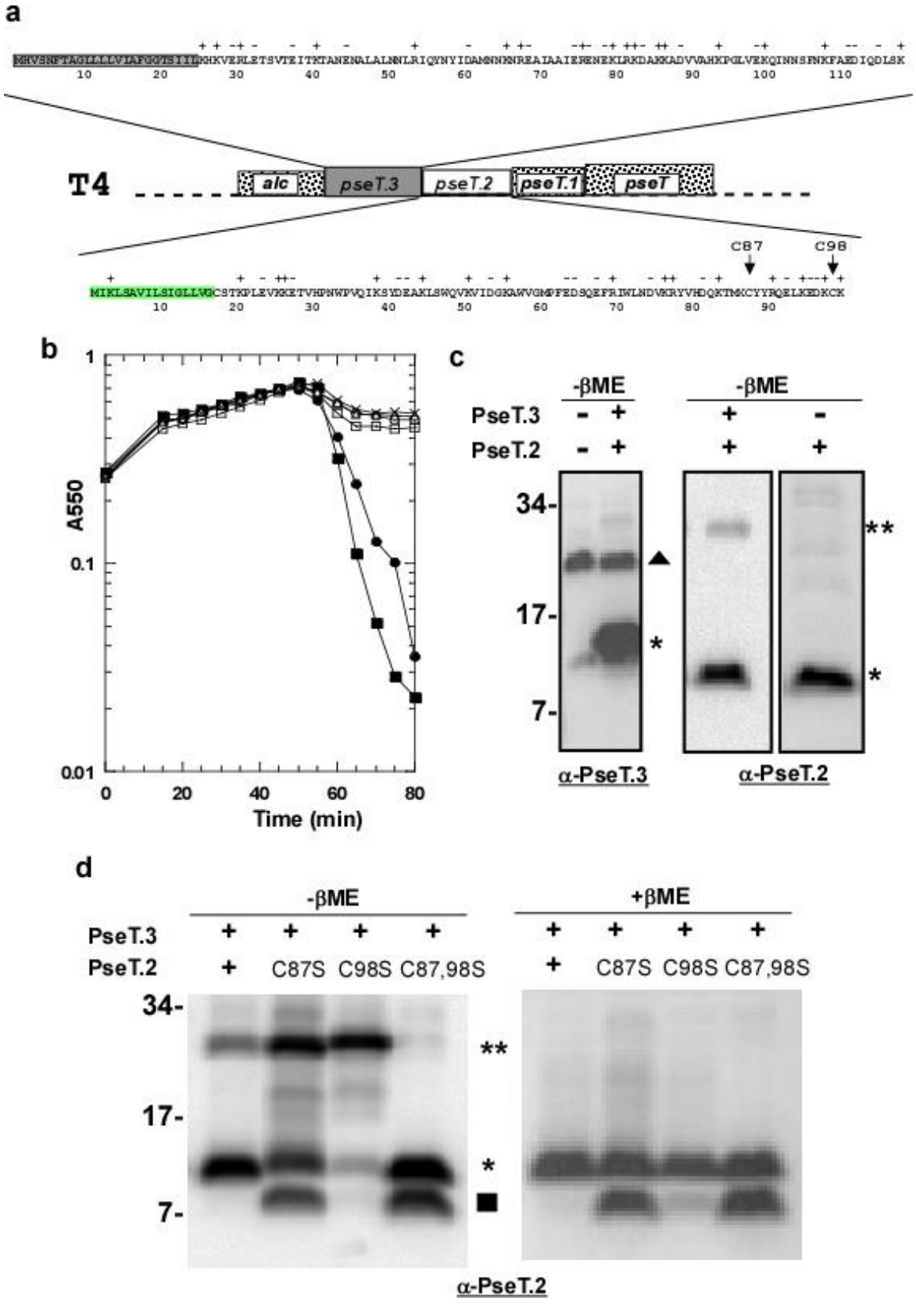
Figure 2.7 Statistical analysis of cysteines in spanins. a) Histogram showing the distribution of periplasmic cysteines in o-spanins w.r.t number of cysteines in their i-spanin counterparts. X-axis indicates the number of cysteines in i-spanins while the Y-axis indicates the number of spanin systems with that combination of periplasmic cysteines in i- and o-spanins. Red, blue, green, purple and yellow indicate the populations with 0,1,2,3, or 4 cysteines in the o-spanin. b) Histogram showing the distance distribution of periplasmic cysteines in spanins closest to the OM (grey bars) and IM (white bars). c) Histogram showing the distance distribution of periplasmic cysteines in spanins; closest cysteine from the OM in o-spanin (yellow bars), closest cysteine from the IM in i-spanin (green bars), closest cysteine from the C-terminal interaction site in o-spanin (blue bars) and closest cysteine from the C-terminal interaction site in i-spanin (red bars). d) Histogram showing the distance distribution of periplasmic cysteines in spanin from the OM (grey bars) and C-terminal interaction site (white bars) in spanin systems with only one periplasmic cysteine.

length of the individual spanins. The proximal cysteine distance to OM in the o-spanin was much less than the distance of the proximal i-spanin cysteine to the IM, with the former peaking ~ 30% of sequence length of the o-spanin, while the latter peaked ~ 60-70% of sequence length of the i-

spanin. To understand whether the proximity of the cysteine to the OM or the C-terminal interaction site was important for function, we analyzed the subset of spanins with only one cysteine and examined the position of the lone cysteine across the periplasmic domain length (Table S2.5e and Fig. 2.7d). In all cases, the lone cysteine was closer to the heterotypic interface than the OM, suggesting that a cysteine closer to the interaction site stabilizes the interaction between the spanin complexes.

To supplement our findings on whether the periplasmic cysteine requirement is general for all two-component spanins, we decided to test if cysteines were required in systems other than lambda. The paradigm phage T4 not only has a spanin pair with a different architecture (separated) but also a different arrangement of periplasmic cysteines; the i-spanin PseT.3 has no cysteines whereas the o-spanin PseT.2 has two cysteines at position 87 and 98 (Fig. 2.8a). First, to confirm the functionality of the T4 spanins, we tested the ability of *pseT.2* and *pseT.3* to complement the spanin lysis defect in a $\lambda SRRz_{am}Rz1_{am}$ background and found that they were functionally equivalent to *Rz-Rz1* (Fig. 2.8b). However, both single and double Cys to Ser substitution alleles of *pseT.2* generated, were found to be lysis-defective (Fig. 2.8b), indicating that, unlike the case in the lambda spanins, both the cysteines in the o-spanin were essential for lysis. When analyzed by western blotting under non-reducing conditions, the T4 i-spanin PseT.3 did not form an SDS-stable dimer *in vivo,* as expected due to the lack of periplasmic cysteine residues (Fig. 2.8c). Unexpectedly, only 30% of the o-spanin PseT.2 was found to be in dimers sensitive to reducing agent (Fig. 2.8c,d). This PseT.2 disulfide-bonded dimer species was detected only in the presence of PseT.3, suggesting a pathway in which PseT.2-PseT.3 complexes form first, with the former carrying the two Cys residues in an intramolecular

Figure 2.8 The role of periplasmic cysteines in T4 spanins. a) Primary structure analysis of the separated spanins *pseT.3* (i-spanin) and *pseT.2* (o-spanin) from T4. Unlike the traditional lysis cassette as in lambda, these genes are not located near the T4 holin (*t*) and endolysin (*e*). The inset shows the predicted primary structures of PseT.3 and PseT.2 with their TMD and signal sequence highlighted in gray and green, respectively. pseT.3 (black) encodes a 117 aa i-spanin and pseT.2 encodes an 83aa o-spanin. The T4 o-spanin has a predicted helix at the C-terminus and is comparatively larger than the o-spanin in lambda which lacks any predicted or detected helical structure. The position of predicted coiled coil structures in the i-spanin are shown by open rectangles. The two periplasmic cysteines in PseT.2, at positions 87 and 98 are shown by arrows. b) Lysis profiles of T4 spanin cysteine mutants. MC4100 (λ*900Rz*$_{am}$*Rz1*$_{am}$) lysogens grown in LB supplemented with 10 mM MgCl2, carrying the following plasmids, were induced at time=0 and growth was monitored at A550: pRE (-X-); p*RzRz1* (-■-); p*pseT.3pseT.2* (-●-); p*pseT.3pseT.2*$_{C87S}$ (-□-); p*pseT.3pseT.2*$_{C98S}$ (-Δ-); p*pseT.3pseT.2*$_{C87,98S}$ (-○-). c,d) Western blot analysis of T4 spanin cysteine mutants: TCA precipitates from induced MC4100 (λ*900Rz*$_{am}$*Rz1*$_{am}$) lysogens carrying the indicated allele were prepared and analyzed in the absence or presence of β-mercaptoethanol as indicated above the gel. For each analysis, the spanin antibody used is indicated at the bottom of each panel. The location of monomeric and dimeric species of PseT.3 and PseT.2 are indicated by (single asterisk) and (double asterisk). Additionally, putative degradation products are indicated by square on the right of each blot. Filled triangles indicate a background band. The alleles are indicated above each lane. Molecular markers in kDa are indicated to the left.


a
T4
alc
pseT.3
pseT.2
pseT.1
pseT
C87
C98
b
A550
Time (min)
c
-βME
-βME
PseT.3
PseT.2
34-
17-
7-
α-PseT.3
α-PseT.2
d
-βME
+βME
PseT.3
PseT.2
C87S
C98S
C87,98S
34-
17-
7-
α-PseT.2

disulfide linkage, followed by an isomerization resulting in the formation of two intermolecular linkages between the cognate Cys residues (Fig. 2.8d). The formation of this doubly-linked complex is required, since the proteins with single intermolecular linkages are non-lytic, despite an enhanced level in both cases (Fig. 2.8d, lanes 2 and 3). Indeed, the cysteines are conserved in those positions throughout the members of the T4 o-spanin family. Thus, the T4 system, although conforming to the general rule of requiring at least one homotypic intermolecular disulfide bond, is more constrained than the lambda system, in terms of requiring two such linkages.

In both cases examined so far, lambda and T4, it is intriguing that the requirement for covalent linkages through the disulfide bonds involves homotypic i-spanin or o-spanin dimers. A requirement for heterotypic i-spanin/o-spanin linkages would not have been surprising, in that it would have provided a covalent bridge between the IM and OM, as is indeed the case for the u-spanins. However, this would require the presence of at least one cysteine in both the spanin components, which is not the case as found in this study. A possible explanation for this observation is that the spanin complexes are subjected to robust forces at the interaction site orthogonal to the axis of the complex, rather than along it. In this scenario, a covalent link via an intermolecular disulfide bond near the junction of spanin interaction in either of the spanin components is necessary to counter these extreme forces. It is possible the intermolecular linkage also stabilizes the collapsed conformation hypothesized for spanin function (Fig. 2.2c).

Approximately 4% of the two-component spanins were found to lack periplasmic cysteines entirely. While sequencing errors can potentially result in false-negatives (spanins with no cysteines) for this analysis, the number seems too high to be attributed to just sequencing quality.

However, none of these phages have been tested for lysis morphology, and thus it would be premature to speculate on how the lack of the covalent linkage is compensated either intragenically or by another as yet unidentified lysis factor.

Diversity of the unimolecular spanins

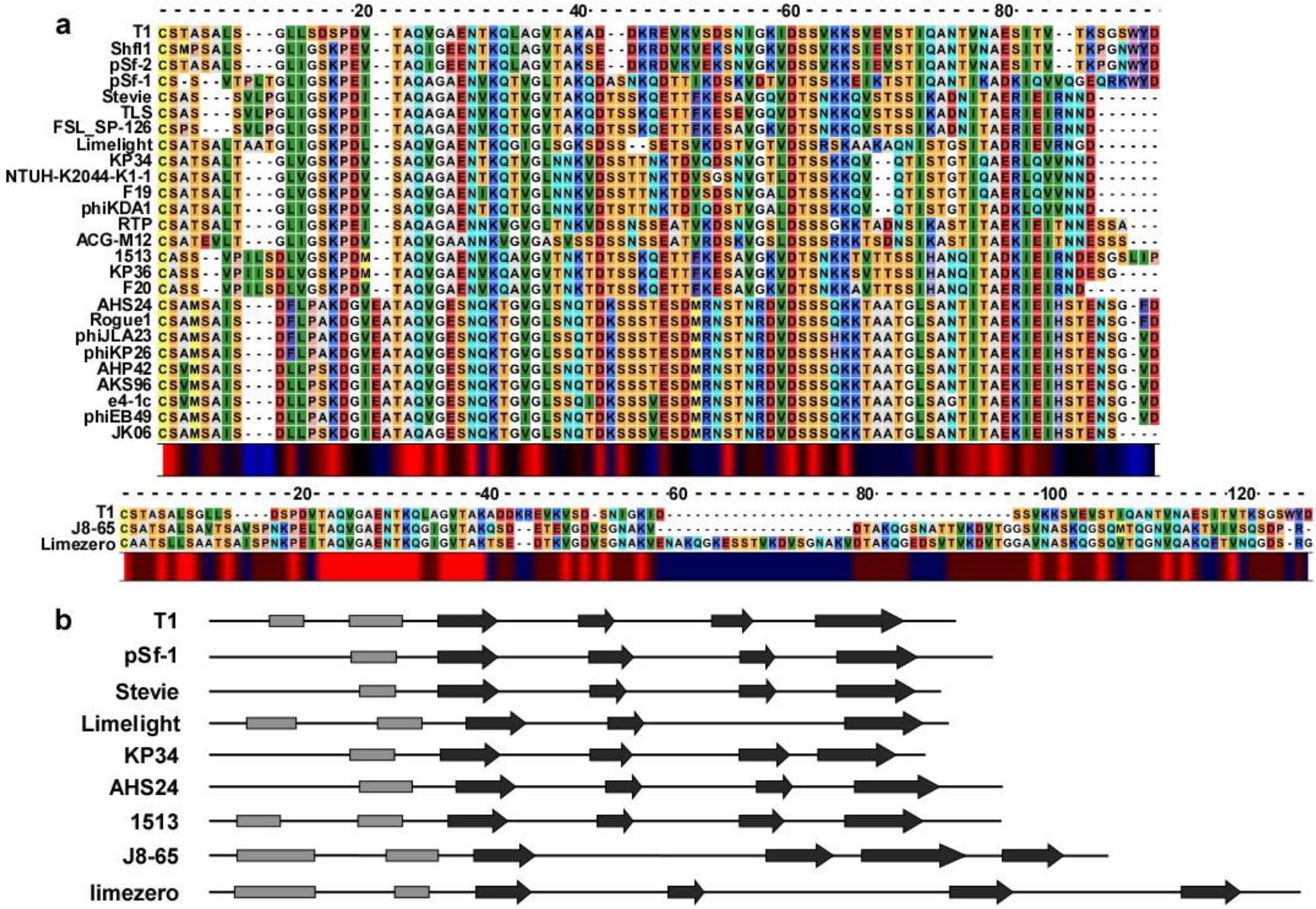
Using the same strategy as used to group two-component spanins into families, the 58 u-spanins were grouped into 13 families, of which 6 were singletons (Table 2.4). Of these 58 u-spanins, there were 3 pairs of identical u-spanin sequences that were identified by grouping sequences that were 100% identical over 100% sequence length (Table 2.5). The family represented by T1gp*11* was the largest with 28 members. Upon aligning the sequences of the members of the T1 family, we noticed various residues spread throughout the periplasmic domain that seem to be involved in predicted secondary structures, were conserved across all members (Fig. 2.9a). Indeed, analysis of predicted secondary structure revealed that members of the gp*11* family, although different in lengths, maintain similar beta sheet secondary structure distribution (Fig. 2.9b). A short alpha helical stretch followed by 4 or more beta sheet elements, connected through flexible linkers, was the dominant pattern of the periplasmic domain of all the family members. Even J8-65 and Limezero, which were unusually long and had the least sequence similarity to other members from the gp*11* family, preserved the secondary structure pattern. The next largest family consisted of 11 members, all from *Pseudomonas* phages represented by PB1. Notably, all of these u-spanins possessed an unusual lipobox motif, AWAC (see next section).

All the members of the T1gp*11* u-spanin family, except for J8-65 and Limezero, were part of lysis cassettes encoding a pinholin and SAR-endolysin, initially suggesting that this type

of lysis system probably favored the u-spanin over a two-component spanin system. However, when we analyzed the lysis cassettes of all the phages encoding u-spanins, of the 46 genomes where we could identify the endolysin, 29 were SAR endolysins and 17 were canonical endolysins (Table S2.1). Thus, it can be concluded that the u-spanins can function with either of the two types of holin-endolysin systems. In contrast to the ubiquitous presence of cysteines in the two-component spanins, none of the 58 u-spanins contained a periplasmic cysteine residue. Thus, we conclude that the constraint in the two-component spanin systems that requires a covalent homodimer linkage near the heterotypic interface does not apply to u-spanin systems.

The diversity and fundamentally different primary and secondary structure characteristics strongly indicate that u-spanins evolved independently of the two-component spanins. One evolutionary path could be that the u-spanins could have originated from an o-spanin, in which a mutation in the stop codon resulting in extension of the C-terminal domain, gaining a C-terminal TMD. However, none of the u-spanins have detectable similarity to any o-spanin, so this scenario remains speculative. There is not yet any physiological or biochemical data addressing how the u-spanins function. Nevertheless, given that both the spanin complex and the u-spanin physically connect the IM and OM, a model for u-spanin function can be proposed on lines, similar to the previously proposed model for two-component spanin function [117]. The most parsimonious idea is that u-spanins also effect OM disruption by causing IM-OM fusion (Fig. 2.2c), albeit through a collapsing conformational change involving beta-sheets rather than coiled-coil helical domains. The differences in secondary structure content and periplasmic cysteines, suggest that the u-spanins, even though achieving the same end result as the two-component spanins, function in a fundamentally different way. Proteins with very different structural

Figure 2.9 The T1gp*11* u-spanin family. a) Sequence alignment of the T1gp*11* family. Shown here is the alignment of the periplasmic domain of members of the T1gp*11* family. The sequences are labelled with the respective phage names on the left. Conservation at different positions is shown below with colors (Red indicates maximum conservation and blue indicates least conservation). A scale indicating the relative position of the residues and the approximate size of various domains is shown above the alignments. b) Secondary structure analysis of the periplasmic domain of T1gp11 family. Shown here is a secondary structure distribution of representative members within the T1gp11 family with varying predictions. Grey rectangles indicate alpha helices and black arrows indicate beta sheets and are drawn to scale, according to their respective lengths.


a
20
40
60
80
T1
Shfl1
pSf-2
pSf-1
Stevie
TLS
FSL_SP-126
Limelight
KP34
NTUH-K2044-K1-1
F19
phiKDA1
RTP
ACG-M12
1513
KP36
F20
AHS24
Rogue1
phiJLA23
phiKP26
AHP42
AKS96
e4-1c
phiEB49
JK06
20
40
60
80
100
120
T1
J8-65
Limezero
b
T1
pSf-1
Stevie
Limelight
KP34
AHS24
1513
J8-65
limezero

composition accomplishing the same result has been seen before in viral membrane fusion proteins [138, 153]. For instance, class I viral membrane fusion proteins predominantly contain alpha helical domains, whereas the class II viral membrane fusion proteins are rich in beta sheets. The two-component spanins can be considered analogous to the class I viral membrane fusion proteins, while the unimolecular spanins can be compared to the class II viral membrane fusion proteins [153].

While genetic analysis of the *Rz/Rz1* [134, 135] has given us clues about the intermediate steps of the membrane fusion process by two-component spanins, mechanistic details of how u-spanins achieve OM disruption are lacking due to the limited experimental work. A thorough genetic analysis of T1*gp11* to find non-functional mutants blocked at various steps in the lytic pathway is needed. Moreover, assuming that the u-spanin pathway terminates in an analogous OM-IM fusion event analogous to that proposed for the two-component spanins, a spheroplast fusion assay similar to the one used for two-component spanins needs to be designed to exploit the mutant collection. Biochemical experiments testing the post translational processing pathway of u-spanins and potential oligomerization of u-spanin molecules at different stages of the proposed model can also be informative.

AWAC, a newly identified lipobox

One of the byproducts of this survey of o-spanins was the identification of novel lipobox sequences. The post-translational processing and sorting of lipoproteins from their nascent form to the mature form has been very well studied [77]. The current consensus lipobox sequence, determined by analyzing the signal sequences of existing lipoproteins over the years is

[LVI][ASTVI][GAS][C] [191, 192]. The consensus sequence for the lipoboxes of o-spanins and u-spanins identified in this work, i.e., the 3 residues before the putative lipoylation cysteine, were plotted using WebLogo 3.3 (Fig. 2.10a,b, Table S2.6) [193]. While most o-spanins still followed the consensus for the lipobox, there were a number of deviations, including AWAC, LNGC, FVGC etc. (Table S2.1). We decided to test one of these unusual lipoboxes, AWAC, as no lipoproteins have been experimentally confirmed to have such a lipobox sequence. Moreover, this lipobox was found only in a family of u-spanins, represented by PB1 (Table 2.4), all of which were associated with the pinholin-SAR endolysin systems. To test whether AWAC is really a lipobox and thus if the u-spanins identified could be functional, we constructed a *gp11* construct where the wild type lipobox LSGC was substituted with AWAC. $gp11_{\mathrm{AWAC}}$ was able to complement the *λRzRz1* defect (Fig. 2.10c), demonstrating that AWAC is a valid lipobox motif.

Spanin-less phages: is there another route to Outer membrane disruption?

In approximately 13% of the genomes analyzed, neither a two-component nor a u-spanin system could be identified (Table 2.7). There are multiple reasons that would explain our inability to identify spanins in these genomes, the first being sequencing errors. For example, the *Escherichia* phage ADB-2 contains a gene, B508_00385, that encodes a protein highly similar to the u-spanin T1gp11 (75% identical and e-value of $2e^{-48}$) but does not have an N-terminal lipoylation signal sequence. The fact that this gene is adjacent to the predicted holin (B508_00375) and endolysin (B508_00380) genes, presumably as part of a lysis cassette, adds to the speculation that it could be a u-spanin, but the reading frame was not detected because of a sequencing error. A more common problem affecting the search for spanin candidates is the quality of annotation of many phage genomes. All the three spanin genes have N-terminal

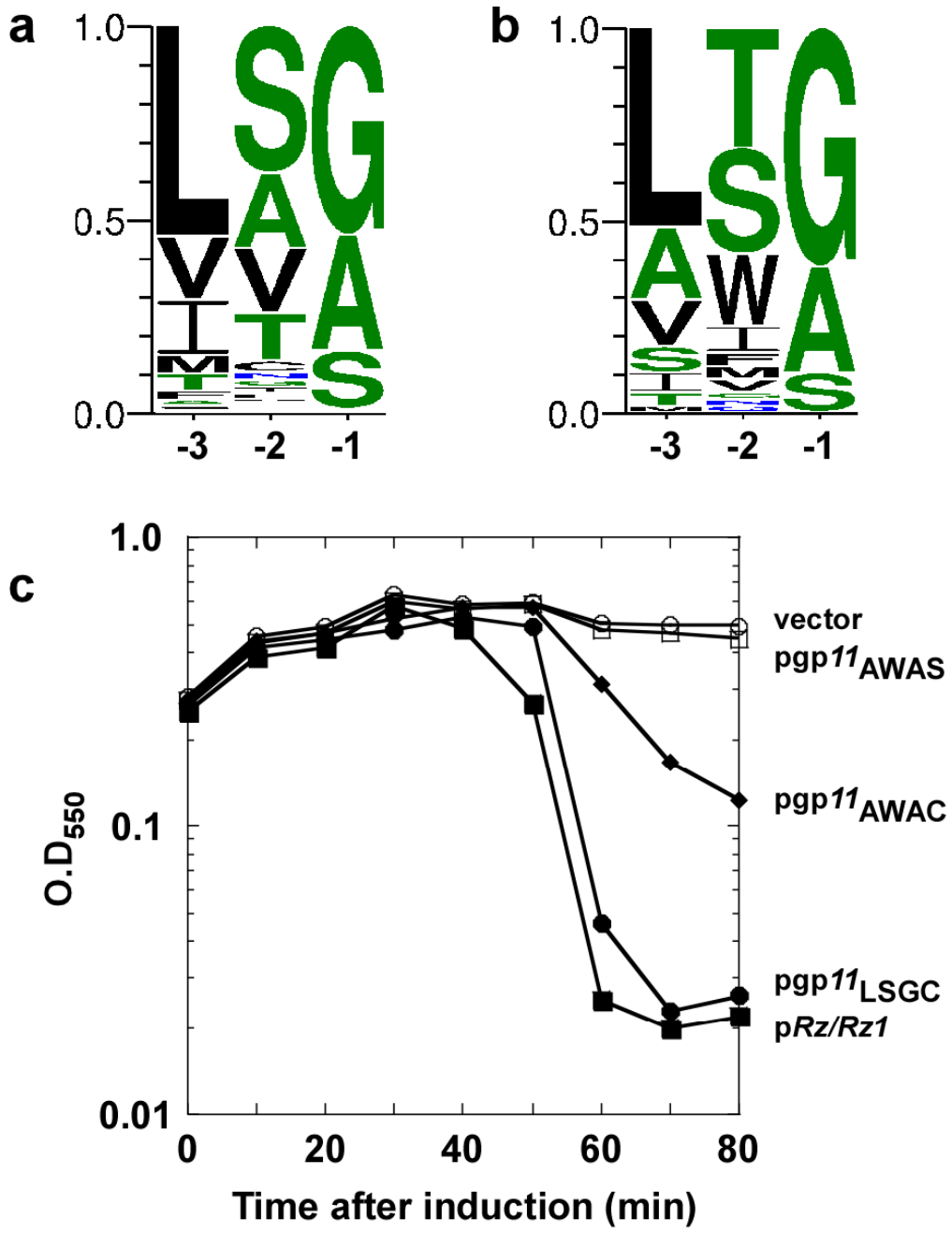


Figure 2.10 A novel lipobox motif. a, b) Consensus lipobox sequences for o-spanins(a) and u-spanins(b). The three residues signify the three residues immediately before the lipoylation cysteine. Lipobox sequences were collected from respective spanins and plotted using WebLogo 3.3. n=528 for a) o-spanins and 58 for b) u-spanins. c) Lysis profiles of gp*11* lipobox variants: MC4100 (*λ900Rz*$_{am}$*Rz1*$_{am}$) lysogens grown in LB supplemented with 10 mM MgCl2, carrying the following plasmids, were induced for lysis at time=0 and growth was monitored as A550: pRE vector (-○-); p*RzRz1* (-●-); pgp*11* (-■-); pgp*11*$_{AWAC}$ (-♦-); pgp*11*$_{AWAS}$ (-□-).

localization signals, either OM lipoprotein or IM N-terminal TMD; recognition of such signals depends absolutely on having the correct initiation codon. However, gene-calling programs often fail at this because of a bias against CDS overlap [194, 195], which is very common in phage genomes. In phage genomes where the lysis genes were clustered, or where BLAST hits to known spanins were found, it was possible for us to manually correct the gene starts. There were 311 cases in which a spanin was either annotated with an incorrect start codon (34 i-spanins, 64 o-spanins and 4 u-spanins) or not annotated at all (13 i-spanins and 196 o-spanins). However, in cases of T4-like phage genomes, lysis genes are not clustered, making it difficult to ascertain whether gene-calling errors resulted in missing spanin genes.

Normally, absent a lysis cluster, searching for spanins is most readily done by first looking for a lipoprotein gene, which are rare in phage genomes and encode the distinctive N-terminal lipobox signal. In 31 genomes, a lipoprotein gene was identified without a closely linked i-spanin candidate (type II integral membrane proteins). For example, Acinetobacter phage vB_AbaM_Acibel004, had gene 41 encoding a lipoprotein, but we could not detect any putative i-spanin gene in the vicinity. Genes encoding type II IM proteins, which could theoretically take up the role of an i-spanin, could be found in other locations. The most extreme case of separation in two-component spanins identified in this study was seen in the phage B40-8, where the predicted i-spanin and o-spanin are separated by ~1kb. However, if there were any phages where the i-spanin and o-spanin were separated by a greater genomic distance, they could not be identified through our current standards.

Other than poor annotation or physical separation of the i-spanin and o-spanin genes, another potential reason for the absence of spanin genes in some phages is that the OM disruption is not necessary for phage lysis in these cases. This could result from differences in

the organization and stability of the cell envelope and/or the natural environment of the host. For example, 20 of the 91 phages with no identifiable spanins were isolated on either *Prochlorococcus* or *Synechococcus* hosts. These marine cyanobacteria are usually found in the euphotic zone in the oceans, where the cells experience extreme physiological conditions [5]. The OM of these hosts might be fragile in those environments, such that the disruption of the IM and PG by the holin and endolysin is sufficient for lysis.

Among the 91 genomes where we could not identify a spanin gene, 34 had no CDS potentially encoding a lipoprotein. This included a number of phages with large genomes, within which it would be impractical to analyze every possible reading frame. Taking 50 kb as an upper limit, we analyzed 16 of the no-spanin genomes with the LipoSearch tool [196], which inspects for potential lipoproteins in every possible reading frame, irrespective of the annotated gene structures. Seven genomes were unambiguously devoid of a lipoprotein coding sequence in any reading frame, even considering potential frame-shift mutations, thus ruling out the presence of either an o-spanin or u-spanin gene. For example, *Acinetobacter* phage Petty has identifiable holin and endolysin genes of the canonical type, but no lipoproteins at all in any reading frame [197]. It can therefore be deduced that these phages employ a different mode of OM disruption. A possible alternative for OM disruption can be derived from recent reports that characterized phage endolysins with membrane penetrating activity [198, 199]. A highly positively charged C-terminal domain with predicted amphipathic helices was a common feature of these membrane-active endolysins, found in ϕKMV-like phages. Our present criteria also do not consider the possibility of proteins with other membrane topologies or anchors, to operate as spanins. For example, a type I (N-out, C-in) IM protein with a C-terminal TMD or an IM lipoprotein could also act as an i-spanin, while an OMP could act as an o-spanin.

Further experiments are needed to understand the mode of OM disruption in these spanin-less phages. A reasonable approach would be to shortlist phages that infect genetically facile hosts like *E. coli, Pseudomonas or Salmonella* and examine their lysis phenotype at the cellular level using phase contrast microscopy. If the phage shows no spanin lysis defect, plasmid libraries containing random segments of the phage genome can be constructed and screened for complementing the spanin lysis defect in the well-established lambda lysogen platform.

Conclusion

For most phages, spanins are required to complete the last step of bacteriophage lysis; i.e. OM disruption. They play a key role in efficiently liberating progeny virions by causing rapid and complete breakdown of the OM, through a pathway that we have proposed to involve IM-OM fusion [117]. In this report, we identified spanins in more than 85% of the phage genomes and added new members to the existing set of two-component and unimolecular spanins. The SpaninDB established here will provide the framework for future spanin identification, classification and characterization. The findings of this research not only provide insights into spanin function, evolution, and domain organization but also the role of intermolecular disulfide bonds in stabilizing the spanin complexes and completing membrane fusion. For phages with no identifiable spanins, hypotheses for other possible modes of OM disruption were suggested. The results suggest several new questions and opens avenues for genetic and biochemical experiments. For example, the mechanism of u-spanin function can be addressed by a spheroplast fusion assay, like the one used for probing the ability of two-component spanins to fuse membranes. Small peptides designed based on the ectodomains of spanins, which could be

potential inhibitors of spanin function, can be titrated into the spheroplast fusion assay to manipulate the timing and efficiency of fusion [200, 201], which can have huge practical implications. Structural studies of spanins would help in understanding the subdomain organization in different spanins and different conformations the spanin complexes assume during different stages of the membrane fusion pathway. They would also shed light on the fundamental question of how u-spanins and two-component spanins approach the same solution even though being totally different in secondary structure compositions. All these suggested future studies using spanins can have serious implications for understanding the cellular envelope of Gram-negative bacteria, and the attributes of OM in particular. Understanding the mechanistic details of spanin function can also pave way for developing biotechnological applications like fusion assays, cargo targeting and delivery systems etc. as well as medical applications such as antimicrobial strategies.

Materials and methods

Bioinformatics procedures

Identification of spanins and implications for automated phage annotation was done as described in Fig. S2.1. All the following information for every phage genome with an identified spanin system was added manually to the SpaninDB; phage name, accession version, architectural class (embedded/overlapped/separated), host, spanin GI (Geneinfo Identifier) number, spanin gene name or coordinates, predicted Shine-Dalgarno sequence for the spanin, primary structure of the spanin, periplasmic domain sequence of the spanin, length of the spanin, number of the periplasmic cysteines in the spanin, lipobox sequence for o-spanin/ u-spanin. All protein sequence and statistical analysis was done using tools on the CPT Galaxy platform (https://cpt.tamu.edu/galaxy/). In brief, FASTA libraries of spanin sequences were created based on architecture and spanin type. These FASTA libraries were used to group the spanins into families using the BLASTCLUST tool. The length coverage threshold and score coverage threshold cutoffs were set to 0.4 and 40 respectively. These parameters were chosen because they gave us the optimal fit between finding distantly related sequences and minimalizing the false positive hits within each family. The families resulting from these parameters were also consistent with the families from our previous report on Rz-Rz1 equivalents [116]. These parameters were changed to 1 and 100 respectively to find identical spanin sequences. Manual corrections to the BLASTCLUST results were done using the online BLASTp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) following the same threshold parameters. Sequence alignments were performed using the desktop version of CLC Main Workbench 7.6.2 (gap open cost=10.0 and gap extension cost=1.0). We used a “non-identical set”, in which only one

sequence was chosen to represent all of its duplicates (Table 2.5), for secondary structure analysis and coiled-coil predictions. Secondary structure analysis was done using Jpred4 [202] (http://www.compbio.dundee.ac.uk/jpred/) and coiled-coils were predicted using Pepcoil with default settings [203] (http://www.bioinformatics.nl/cgi-bin/emboss/pepcoil). The consensus sequence for the lipobox was plotted using the logo generator software WebLogo 3.3 [193] run locally on the CPT Galaxy instance. All data was plotted using the graphing and analysis software Kaleidagraph.

Bacterial strains, culture growth, plasmid constructions and general methods

All cloning procedures were performed using the strain XL1-Blue *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacIqZΔM15 Tn10*] (Stratagene). For the complementation experiments, the temperature sensitive lysogen MC4100 (λ*cI857* Rz_{Q100am}/$Rz1_{W38am}$) [128] carrying the indicated pRE plasmid derivatives were used. The pRE plasmid, a pBR322 derivative, contains the phage λ late promoter pR′ located directly upstream of a multiple cloning site and can be thermally induced. The p*pseT.3pseT.2* plasmid was constructed by replacing the *RzRz1* embedded cassette in p*RzRz1* [115] with the separated *pseT.3* and *pseT.2* genes between the KpnI and BamHI restriction sites. The construction of p*gp11* plasmid used for *gp11* complementation studies has been described earlier [116]. Site directed mutagenesis was used to introduce the cysteine mutations into *pseT.2* and lipobox mutations into *gp11*. Oligonucleotides were purchased from Integrated DNA Technologies, Inc (Coralville, IA). Site directed mutagenesis was confirmed by sequencing at Eton Biosciences. The sequences of the primers used for all PCR amplifications and site-directed mutagenesis are available on request. Cultures were grown in LB medium at 30°C supplemented with appropriate antibiotics, ampicillin (Amp;

100 μg/ml) or kanamycin (Kan; 30 μg/ml) and 10mM $MgCl_2$ and induced for lysis by shifting to 42°C for 15 minutes before shifting to 37°C until the end of the experiment. A_{550} of the cultures was followed using the Gilford Stasar III spectrophotometer.

SDS-PAGE and Western blotting

SDS-PAGE and western blotting were performed as described previously [122]. TCA pellets were washed in two volumes of acetone, resuspended in 1X SDS-PAGE buffer with or without β-mercaptoethanol (100mM BME) and processed as indicated. Protein samples were loaded onto 10% resolving Tris-tricine polyacrylamide gels after loading volumes were normalized according to A_{550} at the time of collection of TCA precipitates. Proteins were transferred to PVDF membrane (Pall Life Sciences) using a Hoefer TE unit at 0.1 mA overnight at 4º C. Antibodies (Sigma Genosys) were generated in rabbits against the synthetic peptides CERENEKLRKDAKKA, corresponding to the PseT.3 residues 74-87 and CWLNDVKRYVHDQKT, corresponding to the PseT.2 residues 71-84. The primary antibodies were used at a dilution of 1:1000 while the secondary antibody, goat-anti-rabbit-HRP (Thermo Scientific), was used at a dilution of 1:5000. Chemiluminescence was detected using a Bio-Rad XR Gel Doc system. SeeBlue Plus2 (Invitrogen) pre-stained standard served as a molecular mass standard.

Tables

Table 2.1 a) Spanin statistics

Type of spanin system	Number of phages
embedded 2CS	182
overlapped 2CS	228
separated 2CS	118
unimolecular spanins	58
No spanins found	91
Total	677

b) Spanin annotation problems

Annotation problem	Number of spanin systems
i-spanin not identified	13
o-spanin not identified	196
i-spanin wrong start	34
o-spanin wrong start	64
i-spanin annotated wrong	28
o-spanin annotated wrong	4

Table 2.2 i-spanin families

Family representative (No. of members)	Members
Embedded	
Lambda (83)	Lambda, H-19B, M6, MP1412, PEp14, YuA, Pollock, FSL_SP-058, FSL_SP-076, 1720a-02, Bcep22, BcepMigl, DC1, Gifsy-2, BcepIL02, ES2, Gifsy-1, Fels-1, ST64b, ES18, FSL_SP-016, HK620, SPN3UB, cdtI, mEp460, spn9CC, vB_SosS_Oslo, BP-4795, HK106, HK633, SE1, ST104, VT2-Sakai, YYZ-2008, mEp235, phiSG1, 933W, ENT39118, Min27, Stx1, Stx2-86, Stx2-II, Stx2-I, Stx2_converting_phage_vB_EcoP_24B, TL-2011c, 21, DE3, Eta, HK629, HK630, pSG3, ST160, ST64T, phi80, SfI, PhiES15, 2851, CUS-3, HK542, HK544, HK75, HK97, Hk022, Sf101, Stx2, c341, epsilon34, mEP234, mEPX1, mEp043_c-1, mEp213, sf6, vB_SemP_Emek, P22, Phi20, Phi75, HK140, HK446, P13374, POCJ13, PS34, mEpX2, phiEt88
T7 (39)	T7, vB_EamP-L1, K11, K30, KP32, gh-1, phiPSA2, Berlin, Yep-phi, Yepe2, YpP-G, phi15, 13a, MmP1, PPpW-4, R, T3, T7M, Y, YpP-R, YpP-Y, YpsP-G, phiA1122, phiIBB-PF7A, phiSG-JL2, phiYeO3-12, vB_YenP_AP5, 285P, BA14, CR8, FE44, IME15, Vi06, K1F, PE3-1, EcoDS1, Phi-S1, Kvp1, CR44b
Bcep176 (15)	Bcep176, PPpW-3, RSK1, phiPSA1, BcepC6B, KS9, eiAU-183, E1, SPC32H, SPC32N, SPN1S, SPN9TCW, TL-2011b, epsilon15, phiV10
N4 (12)	N4, vB_EamP-S6, RG-2014, Bp4, EC1-UPM, JWDelta, ECBP1, pSb-1, vB_EcoP_G7C, JWAlpha, vB_EcoP_PhAPEC5, vB_EcoP_PhAPEC7
Groups of 4	(**B3**, JBD25, JBD18, JBD67), (**KS5,** Smp131, phiRSA1, RSY1)
Groups of 3	(**phiE125**, phi1026b, phi644-2), (**HK225**, ZF40, mEp237), (**LIT1**, vB_PaeP_C2-10_Ab09, Luz7)
Groups of 2	(**BcepMu**, phiE255), (**F10**, vB_PaeP_Tr60_Ab31)
Singletons	phiW-14, Cr30, Xfas53, ECML-117, PY54, N15, BcepB1A, EcP1, S1, F116, phiPLPE, vB_EcoM_ECO1230-10

Table 2.2 (continued)

Family representative (No. of members)	Members
Overlapped	
Jersey* (27)	Jersey, L13, SETP3, SS3e, wksl3, K1G, K1H, K1ind3, SETP7, vB_SenS-Ent1, vB_SenS-Ent2, vB_SenS-Ent3, vB_SenS_AG11, SE2, SETP13, FSL_SP-101, K1ind1, K1ind2, EK99P-1, EP23, HK578, JL1, SSL-2009a, SO-1, FSL_SP-031_SIS, FSL_SP-038_SIS, FSL_SP-049_SIS
DMS3 (16)	DMS3, D3112, F_HA0480sp-Pa1651, JBD24, JBD26, JBD30, JBD5, JBD88A, JD024, LPB1, MP22, MP29, MP38, MP42, MP48, PA1-KOR-2010
JH2 (15)	JH2, ECBP2, KBNP1711, NJ01, EC6, FO1a, FSL_SP-010, FSL_SP-012, FSL_SP-107, Felix01, Moogle, Mushroom, PhiEco32, UAB_phi87, WV8
T5* (10)	T5, EPS7, Stitch, SPC35, bV_EcoS_AKFV33, vB_EcoS_FFH1, DT57C, My1, phiR201_SIS, Shivani_SIS
phiKMV (10)	phiKMV, LKD16, Luz19, MPK6, MPK7, PT2, PT5, phikF77, vB_Pae-TbilisiM32, phi2
KS14 (9)	KS14, ST79, ENT90, Fels-2, RE-2010, phi52237, phiE12-2, phiE202, KL3
F1 (8)	F1, Bk, Fz, Pr, R/C, S708, Tb, Wb
Chi (6)	Chi, FSL_SP-030, FSL_SP-088, FSL_SP-124, iEPS5, Enc34
Groups of 5	(**P2**, Wphi, fiAA91-ss, L-413C, PsP3), (**phiCbk**, CCrKarma, CcrMagneto, CcrRogue, CcrSwift), (**PR3**, PR4, PR772, PRD1, PR5), (**CP1**, OP1, Xop411, Xp10, phiL7)
Groups of 4	(**Bcep1**, Bcep43, Bcep781, BcepNY3), (**CP8**, NCTC12673, CP30A, CPX), (**KL1**, 73, vB_Pae-Kakheti25, vB_PaeS_SCH_Ab26)
Groups of 3	(**PAK_P1**, vB_PaeM_C2-10_Ab1, JG004), (**9NA**, FSL_SP-062, FSL_SP-069), (**phiEa104**, PhiEa21-4, vB_EamM-M7), (**Era103**, phiEA100, phiEa1H), (**K139**, Kappa, VPUSM_8), (**phiHSIC**, Jenny_12G5, pYD38-B), (**MSW-3**, PEi2, vB_AsaM-56[#])
Groups of 2	(**phiJL001**, RDJL_Phi_1), (**phiMHaA1**, vB_MhM_1152AP), (**phi92**, phAPEC8), (**SfIV**, SfV), (**Mu**, D108), (**P1**, P7), (**PAP2,** 119X), (**APSE-1**, APSE-2), (**pIS4-A**, pYD38-A), (**Aaphi23**, S1249), (**HP1**, HP2), (**HK639**, mEP390), (**Paz**, Prado[#]), (**phi1402**, phi1422), (**Bf7**, LKA1[#]), (**DFL12phi1**, EE36P1_SOS[#])*
Singletons	OP2,9g, phiKZ, EL, phiCTX, vB_VpaS_MAR10, 201phi2-1, KPP23, PhiO18P, vB_CsaP_GAP52, KS10, Kpp25, PBC5, Ea35-70, OBP, PAK_P5, PhiPsa374, ENT47670, Xp15, vB_XveM_DIBBI, 7-11, CCrColossus, phiAS7, BcepGomr, phiR8-01, BcepNazgul, Presley, SuMu, phi80-18, vB_CskP_GAP227, RSB3, vB_RleM_PPF1, ESSI-2, Cd1, RSJ2, PY100, RSB1, PM1, AF, vB_EamM-Y2, vB_RleS_L338C, Redjac, phiEcoM-GJ1, vB_RglS_P106B, SSU5, UAB_phi78, phi1M2-2, Salvo

Table 2.2 (continued)

Family representative (No. of members)	Members
Separated	
T4 (40)	T4, Bp7, CC31, JS10, JS98, PG7, ime08, phiR1-RT, vB_YenM_TG1, AR1, ECML-134, PS2, PST, RB14, RB32, RB51, SP18, Shfl2, T4T, e11-2, ime09, pSs-1, phiD1, vB_EcoM-VR20, vB_EcoM-VR7, vB_EcoM_ACG-C40, vB_EcoM_VR25, vB_EcoM_VR26, wV7, RB69, Shf125875, hx01, vB_EcoM_JS09, S16, STML-198, Moon, JSE, Phi1, RB49, vB_EcoM_PhAPEC2
KP27 (7)	KP27, Miller, RB43, KP15, Lw1, RB16, vB_CsaM_GAP161
VP4 (7)	VP4, ICP3_2007_A#, ICP3_2008_A#, ICP3_2009_B#, ICP3#, N4_(Vibrio)#, VP3#
phiAS4 (6)	phiAS4, 25, 31, 44RR2.8t, Aes012, Aes508
Groups of 4	(**KVP40**, nt-1, phi-pp2, VH7D), (**4MG**, Av-05, PVP-SE1, vB_CsaM_GAP31), (**D3**, PAJU2, phi297, vB_PaeS_PMG1)
Groups of 3	(**Aeh1**, PX29, phiAS5), (**CP21**, CP220, CPt10), (**rv5**, 2_JES-2013, vB_EcoM-FV3), (**CR3**, CR9, phiTE), (**K1-5**, K1E#, SP6#)
Groups of 2	(**CC2**, 65), (**PBECO_4**, 121Q)
Singletons	Marshall, B40-8, Pf-WMP3, CR5, phiR1-37, pVp-1, Ac42, vB_CsaM_GAP32, PhiKO2, phi_3, 133, vB_RleM_P10VF, JD001, BcepF1, 1M3-16, Acj9, AH2, phage_7-7-1, PhiP27, F108, Sano

* indicates family with members from different architectures.
# indicates members that were manually added to the group after BLAST analysis for short sequences.
Representative phages of each family are highlighted in **bold.**

Table 2.3 o-spanin families

Family representative (No. of members)	Members
Embedded	
HK97 (32)	HK97, 2851, CUS-3, HK106, HK140, HK446, HK542, HK544, HK633, HK75, Hk022, P13374, P22, POCJ13, PS34, Phi20, Phi75, ST160, ST64T, Sf101, Stx2, c341, epsilon34, mEP234, mEPX1, mEp043_c-1, mEp213, mEp235, mEpX2, phi80, sf6, vB_SemP_Emek
Lambda (28)	Lambda, PhiES15, phiSG1, phiEt88, pSG3, SfI, ES18, Fels-1, Gifsy-2, ST64b, 1720a-02, 933W, BP-4795, H-19B, Min27, Stx1, Stx2-86, Stx2-II, Stx2-I, Stx2_converting_phage_vB_EcoP_24B, TL-2011c, VT2-Sakai, YYZ-2008, cdtI, 21, DE3, HK629, HK630
Bcep176 (15)	Bcep176, RSK1, SPC32H, SPC32N, SPN1S, SPN9TCW, TL-2011b, epsilon15, phiV10, BcepC6B, KS9, eiAU-183, PPpW-3, E1, phiPSA1
HK620 (14)	HK620, PEp14, FSL_SP-058, FSL_SP-076, Pollock, Gifsy-1, FSL_SP-016, SE1, SPN3UB, ST104, mEp460, spn9CC, vB_SOSS_Oslo, ENT39118
K11 (14)	K11, K30, KP32, MmP1[#], vB_EamP-L1[#], IME15[#], phiSG-JL2[#], phiYeO3-12[#], vB_YenP_AP5[#], EcoDS1[#], K1F[#], PE3-1[#], CR44b[#], CR8[#])
N4 (12)	N4, vB_EamP-S6, Bp4, EC1-UPM, JWAlpha, JWDelta, ECBP1, RG-2014, pSb-1, vB_EcoP_G7C, vB_EcoP_PhAPEC5, vB_EcoP_PhAPEC7
T7 (11)	T7, 13a, R, T3, T7M, Vi06, Y, YpP-R, YpP-Y, YpsP-G, phiA1122
Kvp1 (8)	285P, BA14, Berlin, FE44, Kvp1, Yep-phi, Yepe2, YpP-G
Groups of 4	(**B3**, JBD18, JBD67, JBD25), (**KS5**, Smp131, RSY1, phiRSA1), (**phi15**, PPpW-4, Phi-S1, phiIBB-PF7A)
Groups of 3	(**LIT1**, Luz7, vB_PaeP_C2-10_Ab09), (**phiE125**, phi1026b, phi644-2), (**M6**, MP1412, YuA), (**HK225**, ZF40, mEp237), (**BcepIL02**, BcepMigl, ES2[#])
Groups of 2	(**BcepMu**, phiE255), (**Bcep22**, DC1), (**F10**, vB_PaeP_Tr60_Ab31), (**phiPSA2**, gh-1)
Singletons	vB_EcoM_ECO1230-10, N15, Ecp1, ECML-117, PY54, BcepB1A, F116, S1, phiW-14, Cr30, Xfas53, Eta, phiPLPE

Table 2.3 (continued)

Family representative (No. of members)	Members
Overlapped	
Jersey* (27)	Jersey, FSL_SP-101, K1G, K1H, K1ind1, K1ind2, K1ind3, L13, SE2, SETP13, SETP3, SETP7, SS3e, vB_SenS-Ent1, vB_SenS-Ent2, vB_SenS-Ent3, vB_SenS_AG11, wksl3, FSL_SP-031_SOS, FSL_SP-038_SOS, FSL_SP-049_SOS, EK99P-1, EP23, HK578, JL1, SO-1, SSL-2009a
DMS3 (16)	DMS3, D3112, F_HA0480sp-Pa1651, JBD24, JBD26, JBD30, JBD5, JBD88A, JD024, LPB1, MP22, MP29, MP38, MP42, MP48, PA1-KOR-2010
P2 (15)	P2, ENT90, ST79, Fels-2, L-413C, PsP3, RE-2010, Wphi, fiAA91-ss, KL3, KS14, phi52237, phiE12-2, phiE202, phiCTX
JH2 (11)	JH2, EC6, FO1a, FSL_SP-010, FSL_SP-012, FSL_SP-107, Felix01, Moogle, Mushroom, UAB_phi87, WV8
T5* (10)	T5, My1, DT57C, EPS7, SPC35, Stitch, bV_EcoS_AKFV33, vB_EcoS_FFH1, Shivani_SOS, phiR201_SOS
BcepNazgul* (10)	BcepNazgul, 73, vB_Pae-Kakheti25, vB_PaeS_SCH_Ab26, KL1, HK639, mEP390, PY100, AF, AH2_SOS
phiKMV (9)	phiKMV, LKD16, Luz19, MPK6, MPK7, PT2, PT5, phikF77, vB_Pae-TbilisiM32
Chi (8)	Chi, SSU5, Enc34, FSL_SP-030, FSL_SP-088, FSL_SP-124, iEPS5, Redjac
F1 (8)	F1, Bk, Fz, Pr, R/C, S708, Tb, Wb
Mu (7)	Mu, D108, ENT47670, SfIV, SfV, APSE-1, APSE-2
CP1 (7)	CP1, phiL7, OP1, Xop411, Xp10, Prado, Paz
phiCbk (6)	phiCbk, CCrKarma, CcrMagneto, CcrSwift, CCrColossus, CcrRogue
PR3 (5)	PR3(tecti), PR4(tecti), PR5(tecti), PR772(tecti), PRD1(tecti)
Groups of 4	(**Bcep1**, Bcep43, Bcep781, BcepNY3), (**CP8**, CP30A, CPX, NCTC12673), (**PAK_P1**, JG004, PAK_P5, vB_PaeM_C2-10_Ab1), (**9NA**, ESSI-2, FSL_SP-062, FSL_SP-069), (**NJ01**, ECBP2, KBNP1711, PhiEco32), (**MSW-3**, PEi2, vB_AsaM-56, JD001_SOS)*
Groups of 3	(**K139**, Kappa, VPUSM_8), (**phiEa104**, PhiEa21-4, vB_EamM-M7), (**Era103**, phiEA100, phiEa1H), (**phiHSIC** , Jenny_12G5, pYD38-B)
Groups of 2	(**KPP23,** RDJL_Phi_1), (**phi1402**, phi1422), (**phi92**, phAPEC8), (**phiAS7**, phiR8-01), (**phiMHaA1**, vB_MhM_1152AP), (**HP1**, HP2), (**P1**, P7), (**Aaphi23**, S1249), (**pIS4-A**, pYD38-A), (**PAP2**, 119X), (**PM1**, phiEcoM-GJ1), (**RSB1**, Cd1[#]) (**Bf7**, phi2, LKA1[#]), (**SuMu**, vB_EamM-Y2[#]), (**DFL12phi1**, EE36P1_SOS)*, (**Salvo**, Sano_SOS)*

Table 2.3 (continued)

Family representative (No. of members)	Members
Overlapped	
Singletons	Ea35-70, OBP, EL, Xp15, phiKZ, 201phi2-1, phiJL001, 9g, vB_XveM_DIBBI, KS10, PBC5, Kpp25, OP2, phi80-18, vB_RleM_PPF1, BcepGomr, vB_CsaP_GAP52, 7-11, vB_CskP_GAP227, PhiO18P, Presley, RSB3, vB_VpaS_MAR10, RSJ2, vB_RleS_L338C, vB_RglS_P106B, PhiPsa374, phi1M2-2

Table 2.3 (continued)

Family representative (No. of members)	Members
Separated	
T4 (47)	T4, phiAS5, Ac42, Moon, CC2, CC31, PG7, 65, AR1, ECML-134, JSE, PS2, PST, Phi1, RB14, RB32, RB49, RB51, Shfl2, T4T, e11-2, ime09, pSs-1, phiD1, vB_EcoM_ACG-C40, wV7, PX29, phiR1-RT, vB_EcoM-VR7, vB_EcoM_VR25, vB_YenM_TG1, 133, Aeh1, Bp7, JS10, JS98, RB69, S16, SP18, STML-198, Shf125875, hx01, ime08, vB_EcoM-VR20, vB_EcoM_JS09, vB_EcoM_PhAPEC2, vB_EcoM_VR26
KP27 (7)	KP27, Lw1, Miller, RB16, RB43, KP15, vB_CsaM_GAP161
rv5 (7)	rv5, 2_JES-2013, Av-05, vB_EcoM-FV3, 4MG, PVP-SE1, vB_CsaM_GAP31
VP4 (7)	VP4, ICP3_2007_A, ICP3_2008_A, ICP3_2009_B, ICP3, N4_(Vibrio), VP3
phiAS4 (6)	phiAS4, 25, 31, 44RR2.8t, Aes012, Aes508
Groups of 4	(**KVP40**, nt-1, VH7D, phi-pp2), (**D3**, PAJU2, phi297, vB_PaeS_PMG1), (**K1-5**, K1E, SP6, UAB_phi78_OOS)*
Groups of 3	(**CR3**, CR9, phiTE), (**PBECO_4**, 121Q, vB_CsaM_GAP32), (**CP21**, CP220, CPt10)
pVp-1 (2)	pVp-1, phi_3
Singletons	CR5, B40-8, phiR1-37, Marshall, vB_RleM_P10VF, Acj9, 1M3-16, phage_7-7-1, PhiKO2, BcepF1, Pf-WMP3, PhiP27, F108

* indicates family with members from different architectures.
[#] indicates members that were manually added to the group after BLAST analysis for short sequences.
Representative phages of each family are highlighted in **bold.**

Table 2.4 u-spanin families

Family representative (No. of members)	Members
T1 (28)	Limezero, J8-65, 1513, bV_EcoS_AHP42, bV_EcoS_AHS24, bV_EcoS_AKS96, e4-1c, phiEB49, phiJLA23, phiKP26, vB_EcoS_Rogue1, pSf-1, JK06, KP36, RTP, vB_EcoS_ACG-M12, Shfl1, T1, pSf-2, LIMELIGHT, F20, FSL_SP-126, Stevie, TLS, F19, KP34, NTUH-K2044-K1-1, phiKDA1
PB1 (11)	PB1, 14-1, F8, SPM-1, JG024, KPP12, LBL3, NH4, SN, lma2, vB_PaeM_PAO1Ab27
phi13:2 (4)	phi13:2, phi18:3, phi19:3, phi47:1
phi12:1 (3)	phi12:1, phi17:1, phi18:1
Groups of 2	(**phi10:1**, phi19:1), (**VpV262**, CW02), (**SIO1**, P12053L)
Singletons	RAP44, 11b, CHOED, FCL-2, 9A, ICP2_2013_A_Haiti

Representative phages of each family are highlighted in **bold.**

Table 2.5 Identical spanin families Phages representing their identical partners in the non-identical set are highlighted in bold (See methods).

Family representative (No. of identical members)	**Members**
Embedded	
T3 (7)	T3, R, T7M, Y, YpP-R, YpP-Y, phiA1122
Groups of 4	(**SPC32H**, SPC32N, SPN1S, SPN9TCW), (**933W**, Stx1, Stx2-II, Stx2_converting_phage_vB_EcoP_24B), (**Lambda**, DE3, HK629, HK630)
P22 (3)	P22, Phi20, Phi75
Groups of 2	(**phiE125,** phi1026b), (**SE1**, ST104), (**Stx2-I**, Stx2-86), (**ST160**, ST64T), (**Stx2**, 2851), (**HK75**, HK544), (**mEp213**, mEp043_c-1), (**Yepe2**, YpP-G), (**BA14**, FE44)
Overlapped	
F1 (8)	F1, Bk, Fz, Pr, R-C, S708, Tb, Wb,
JBD5 (5)	JBD5, F_HA0480sp-Pa1651, MP29, MP48, PA1-KOR-2010,
Groups of 3	(**DMS3**, JBD30, MP42), (**MP22**, JBD88A, MP38), (**vB_SenS-Ent1**, vB_SenS-Ent2, vB_SenS-Ent3), (**Era103**, phiEA100, phiEa1H), (**Felix01**, FO1a, UAB_phi87), (**phiKMV**, PT5, vB_Pae-TbilisiM32)
Groups of 2	(**K139**, Kappa), (**phiE12-2**, phi52237), (**phiMHaA1**, vB_MhM_1152AP), (**Bcep43**, Bcep781), (**P1**, P7), (**SS3e**, wksl3), (**phiEa104**, PhiEa21-4), (**FSL_SP-062**, FSL_SP-069), (**K1ind1,** K1ind2), (**pIS4-A**, pYD38-A), (**FSL_SP-010**, FSL_SP-012), (**Aaphi23**, S1249), (**Chi**, FSL_SP-030)
Separated	
ICP3 (5)	ICP3, ICP3_2007_A, ICP3_2008_A, ICP3_2009_B, N4_(Vibrio),
Shfl2 (4)	Shfl2, PST, ime09, pSs-1,
FSL_SP-031 (3)	FSL_SP-031, FSL_SP-038, FSL_SP-049,
Groups of 2	(**PBECO_4**, 121Q), (**AR1**, wV7), (**ECML-134**, vB_EcoM_ACG-C40), (**Miller**, RB43), (**KP27**, KP15), (**Aes012**, 25), (**phiAS4**, Aes508), (**VP4**, VP3)
Unimolecular	
Groups of 2	(**bV_EcoS_AHS24**, vB_EcoS_Rogue1), (**F8**, SPM-1), (**phi13:2**, phi19:3)

Table 2.6 a) Statistics of cysteines in two-component spanins

No. of total cysteines	No. of spanins
0	21
1	72
2	191
3	130
4	87
5	8
6	16
7	0
8	1
Total	526

b) Statistics of cysteines in i-spanins

No. of cysteines in i-spanin	No, of spanins
0	169
1	164
2	179
3	9
4	4
5	0
6	1
Total	526

c) Statistics of cysteines in o-spanins

No. of cysteines in o-spanin	No. of spanins
0	124
1	118
2	240
3	9
4	35
Total	526

Table 2.7 Phages with no identifiable spanins

Phage Name	Host
AB3, Petty, Abp1, YMC-13-01-C62, vB_AbaM_Acibel004, ZZ1	*Acinetobacter*
pAh6-C	*Aeromonas*
vB_AmaP_AD45-P1	*Alteromonas*
A-4L	*Anabaena*
B124-14	*Bacteroides*
BIP-1, BMP-1, Bpp-1	*Bordatella*
P2559S,P2559Y	*Croceibacter*
Pf-WMP4, PSS2, P-SSP7	*Cyanobacteria*
RC-2014	*Dickeya*
phiKT, ADB-2, Phax-I, vB_EcoM_CBA120,ECML-4	*Escherichia*
SH1	*Haloarcula*
JM-2012	*Halocynthia*
phiHAP-1	*Halomonas*
HF2	*Halorubrum*
1961P, KHP30, KHP40	*Helicobacter*
SC1	*Liberibacter*
psiM2	*Methanobacterium*
PhiCh1	*Natrialba*
P-HM1, P-HM2, pRSM4, P-SSM2, P-SSM4, PSSM-7, Syn1, Syn33	*Prochlorococcus*
PM2, h105/1	*Pseudo-alteromonas*
KPP10, lu11, Luz24, PA11, PaP3, Tf, TL	*Pseudomonas*
RSB2, RSL-1	*Ralstonia*
16-3	*Rhizobium*
RcapMu	*Rhodobacter*
Rm378	*Rhodothermus*
vB_SalM_SJ2, Maynard, Vi01, SFP10, vB_SalM_SJ3	*Salmonella*
Spp001, 1/4	*Shewanella*
phiSboM-AG3	*Shigella*
P60, S_CBS4, S-PM2, SRS-M4, S-ShM2, S-SM1, S-SM2, S-SSM5, S-SSM7, Syn19, Syn5, Syn9	*Synechococcus*
BA3	*Thalassomonas*
IN93, P23-45, P23-77, P74-26, phiYS40	*Thermus*
cp-t, phi-A318, SHOU24, SIO-2, vB_VchM-138, VP2, VP5, VP882, VP93	*Vibrio*

CHAPTER III

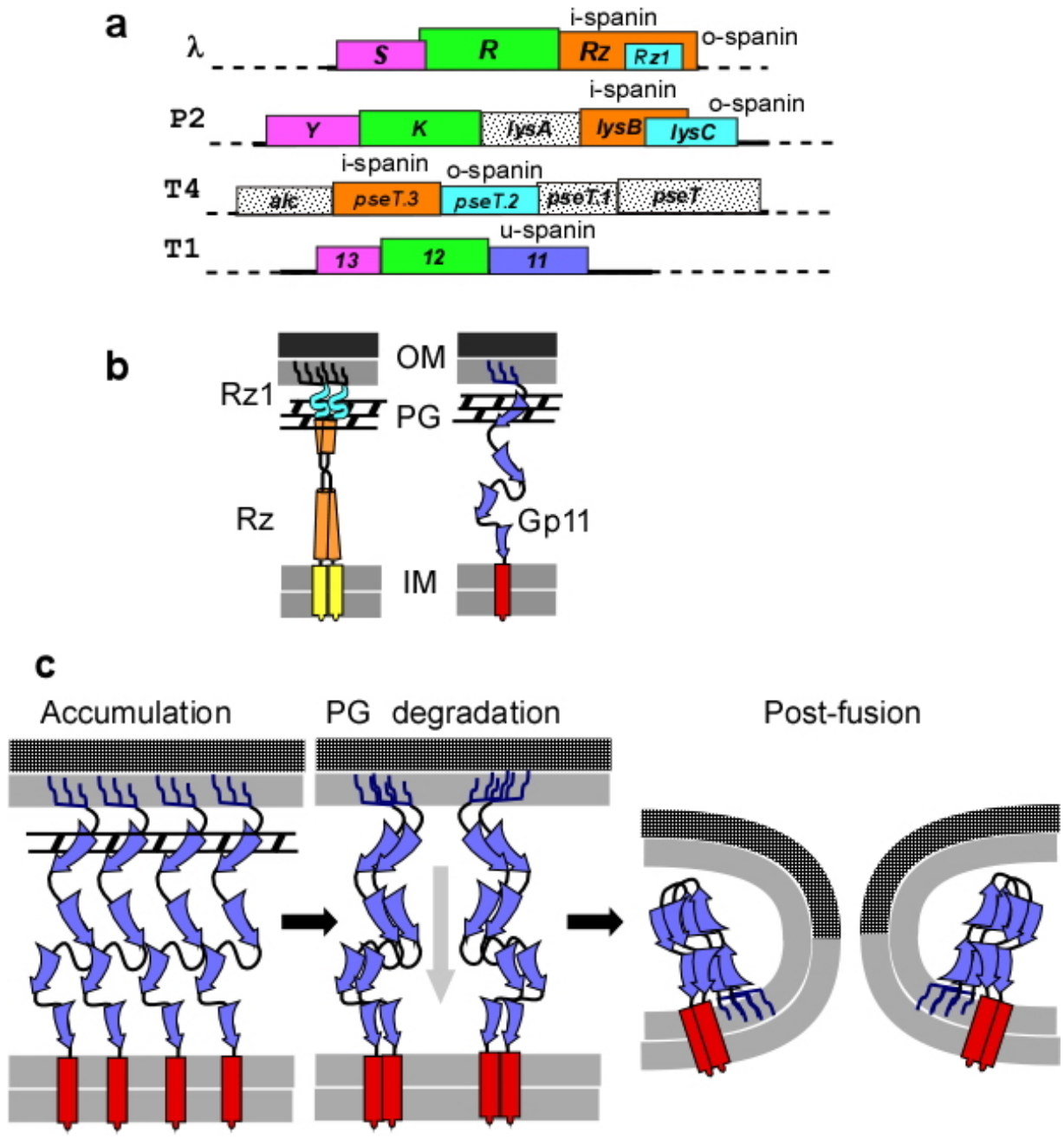
LOCALIZATION AND REGULATION OF THE T1 UNIMOLECULAR SPANIN

Introduction

It has been widely accepted that the *Caudovirales*, which dominate the microbial virome, use the holin-endolysin model for host lysis, in which the holin controls the timing of lysis by controlling access of the muralytic endolysin to the peptidoglycan (PG) [101]. Only recently has it been shown that, in Gram-negative hosts, active disruption of the outer membrane (OM) is also required [114, 128]. In most phages, OM disruption is caused by the spanin complex, consisting of two subunits, an inner membrane (IM) protein (i-spanin) with type II topology and an OM lipoprotein (o-spanin). In the well-studied phage lambda system, the i-spanin Rz and o-spanin Rz1, are encoded by nested genes in the lambda lysis cassette (Fig. 3.1a) [128]. Both spanin subunits accumulate as covalent homodimers linked by disulfide bonds [129]; these homodimers form a heterotetrameric protein bridge through the PG and across the periplasm by virtue of heterotypic C-terminal i-spanin/o-spanin interactions (Fig. 3.1b) [52, 115]. Proposed structural models of these complexes rely heavily on detailed genetic and biochemical analysis [52, 115, 129, 134], focused on the predominantly alpha-helical, coiled-coil structure of the i-spanin. Moreover, there is compelling evidence that the spanin complexes disrupt the OM by causing IM-OM fusion and that this topological function is blocked as long as the PG is intact [117].

Two-component spanins analogous to the Rz-Rz1 proteins can be identified in the genomes of most phages of Gram-negative hosts [116]. However, some phages, including the paradigm phage T1, lack recognizable two-component spanins and instead have a single gene, like *gp11,* in the position immediately distal to the endolysin gene in the lysis cassette (Fig.3.1a).

Figure 3.1 Two component and unimolecular spanins. a) Different spanin genetic architectures from the phages λ (embedded), P2 (overlap), T4 (separate) and T1 (unimolecular). The boxes color coded by pink, green, orange, cyan and purple represent the holin, endolysin, i-spanin, o-spanin and u-spanin respectively in each phage, while the dotted boxes represent genes of unrelated or unknown function. All genes are drawn to scale and labeled with appropriate gene names. The spanin genes are also labeled with their spanin type on top of the gene in each case. b) Cartoon representation of the topology of two-component spanin prototype from phage λ and the u-spanin from phage T1. In λ, i-spanin Rz (orange) is embedded to the IM by an N-terminal TMD (yellow rectangle) and has a periplasmic domain that constitutes two alpha helices (cylinders) connected by a linker, predicted to form coiled coils. The mature form of o-spanin Rz1 (cyan) is attached to the inner leaflet of the OM via three fatty acyl chains (dark blue lines) and has a periplasmic domain predicted to be unstructured. They interact through their C-termini to form the spanin complex, linking the IM and OM through the PG meshwork. In T1, the u-spanin gp*11* (purple) is attached to the inner leaflet of the OM, by the three fatty acyl chains (dark blue lines) at the N-terminus and to the inner membrane, through the C-terminal TMD (red rectangle). The periplasmic domain of gp*11*, predicted to be mainly extended beta sheets (arrows), connects the IM and OM through the PG meshwork. c) Model for gp*11* function. gp*11* accumulates in both the membranes of Gram-negative bacteria stuck within the PG meshwork. Once the PG is degraded by the endolysin, the gp*11* complexes can undergo higher order oligomerization and/or conformational changes along the periplasmic domain (indicated by grey arrow), fusing both the membranes together.



Mature gp*11* is predicted to have an unprecedented localization. It has signals for localization to both membranes; an OM lipoprotein signal and a C-terminal transmembrane domain (TMD) (Fig. 3.1b, 3.2a). This architecture, combined with the ability of gp*11* to complement the lysis defect of $\lambda Rz_{am}Rz1_{am}$ [116], defined gp*11* as the prototype unimolecular spanin (u-spanin).

Unlike the two-component spanins, gp*11* has neither predicted helical structure nor any periplasmic cysteines for disulfide-linked dimerization. Instead, the periplasmic domain of gp*11* is predicted to be dominated by beta strands (Fig. 3.2a). Nonetheless, the obvious analogy between the single polypeptide bridge between the OM and the IM supplied by the u-spanin and the non-covalent complexes spanning the periplasm supplied by Rz-Rz1, suggests the u-spanin also functions by IM-OM fusion (Fig. 3.1c). The differences between the predicted secondary structure of the gp*11* periplasmic domain and the dominant coiled-coil structure of the Rz-Rz1 complex strongly suggest that the fusion pathways are dramatically different, yet functionally equivalent. Here, results of genetic and molecular analysis of the subcellular localization, function and regulation of T1gp*11* are presented and discussed.

Results

Localization signal mutants

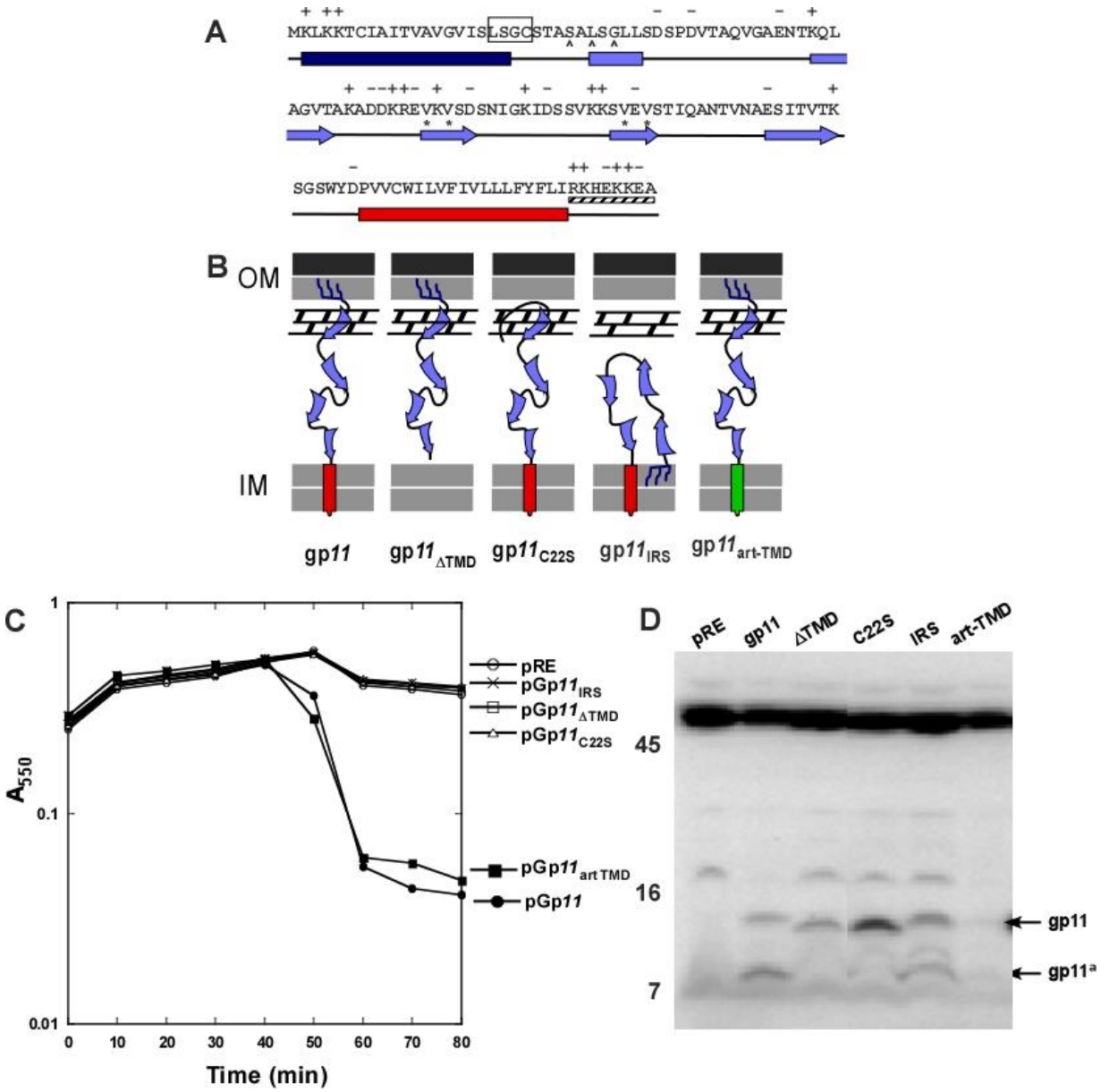
The primary structure of gp*11*, with its predicted N-terminal lipobox and OM-lipoprotein sorting signals and C-terminal TMD, suggests that it is localized to both membranes of the *E. coli* envelope. To test whether predicted localization signals were functional and required, we constructed a set of alleles with altered topological signals (Fig. 3.2b) and tested them for complementation of the lysis-defective $\lambda Rz_{am}Rz1_{am}$ mutant. One of these, the gp$11_{\Delta TMD}$, allele, has the C-terminal TMD deleted and would thus be predicted to localize exclusively to the OM as a mature lipoprotein. Two other alleles encode products predicted to localize only in the IM: gp$11_{C22S,}$ in which the Cys residue that is modified during lipoprotein processing has been replaced by a non-modifiable Ser residue; and gp11_{IRS} (Inner membrane Retention Signal), an

allele in which the residues immediately distal to the processed Cys residue are replaced by Asp residues. This latter mutation is predicted to abolish the OM sorting signal recognized by the Lol system [204], locking the N-terminal lipoylated end in the IM. All these alleles exhibited absolute-defective lysis phenotypes, as judged by the inability to complement the lambda spanin lysis defect (Fig. 3. 2c). To assess whether the C-terminal TMD in function of gp*11* involved specific interactions, we constructed gp$11_{\text{art-TMD}}$, in which the entire TMD of gp*11* is exchanged with an artificial TMD (VLLIIVVVVVVVVIILLI). This TMD substitution allele retained full lysis function (Fig. 3.2c). Taken together, these results demonstrate that both the membrane localization signals of gp*11* are necessary for lytic function and the role of the C-terminal TMD is to anchor gp*11* in the IM, rather than having a required interaction with itself or another integral membrane protein.

To confirm the loss of function was not due to defects in expression, we performed western blotting on whole-cell samples from MC4100 $\lambda Rz_{am}Rz1_{am}$ cultures expressing these constructs in trans. While gp$11_{\Delta\text{TMD}}$, gp11_{C22S} and gp11_{IRS} accumulated at similar level as gp*11*WT, we could not detect any gp$11_{\text{art-TMD}}$ in the blot (Fig. 3.2d). In the case of gp$11_{\text{art-TMD}}$ construct, the TMD substitution is adjacent to the C-terminal cytoplasmic epitope, indicating the lack of immunoblot signal is likely due to distortion of the epitope.

Interestingly, instead of an expected larger unprocessed gp*11* species accumulating in the gp11_{C22S} sample, a smaller product is generated. Sequence analysis with SignalP [205] and LipoP [206] both indicate gp*11* has potential SPI cleavage sites at positions 26, 28 and 30, the 28^{th} position having the highest probability of cleavage (Fig. 3.3a). Thus, we hypothesized that in the absence of SPaseII processing, SPaseI can process the gp*11* precursor to a periplasmic intermediate. To test this idea, the dependence of gp*11* processing on SPaseII was examined by

Figure 3.2 Topological mutants of gp*11*. a) Primary structure of T1gp*11*. Shown here is the primary structure, charge distribution and secondary structure of the T1 u-spanin gp*11*. Dark blue rectangle = N-terminal lipoylation signal sequence; boxed residues = lipobox; purple rectangle = alpha helicle; purple arrows = extended beta sheets; red rectangle = C-terminal TMD. Asterisks denote the alternative start sites and ∧s indicate the potential SPaseI processing sites as predicted by LipoP 1.0. The C-terminal epitope where the gp*11* antibody binds is highlighted by a hatched bar below. b) Cartoon representation of topological mutants of gp*11*. Upon post translational processing, mature gp*11* is translocated to both IM and OM. gp$11_{\Delta TMD}$ is localized to the OM only whereas gp11_{C22S} and gp11_{IRS} all localize to the IM only. gp$11_{art\text{-}TMD}$ is translocated to both the membranes but has a different TMD (green rectangle) compared to the WT gp*11* (red rectangle). c) Lysis phenotype of gp*11* topological mutants. Cultures of MC4100 $\lambda Rz_{am}Rz1_{am}$ carrying pRE plasmids carrying different gp*11* alleles were grown at 30°C until A_{550} was ~0.25. The prophage and the plasmid carrying the spanin were induced by transferring to 42°C at t=0 min and moved to 37°C at t=15 min. The optical density of the culture was followed over time every 10 min. Open circles = no insert (pRE vector only control), closed circles = gp*11*, open square = gp$11_{\Delta TMD}$, open triangles = gp11_{C22S}, crosses = gp11_{IRS}, closed squares = gp$11_{art\text{-}TMD}$. The slight decrease in absorbance of the cultures which did not lyse, is due to the change of cellular shape from rectangle to sphere. d) TCA precipitates were collected from cultures above at 40 min after induction and analyzed by western blotting. The samples in each lane are labelled at the top. The bands corresponding to the mature gp*11* and gp11^{a}, the smaller molecular weight product due to translation from an alternate start site, are indicated by arrows.



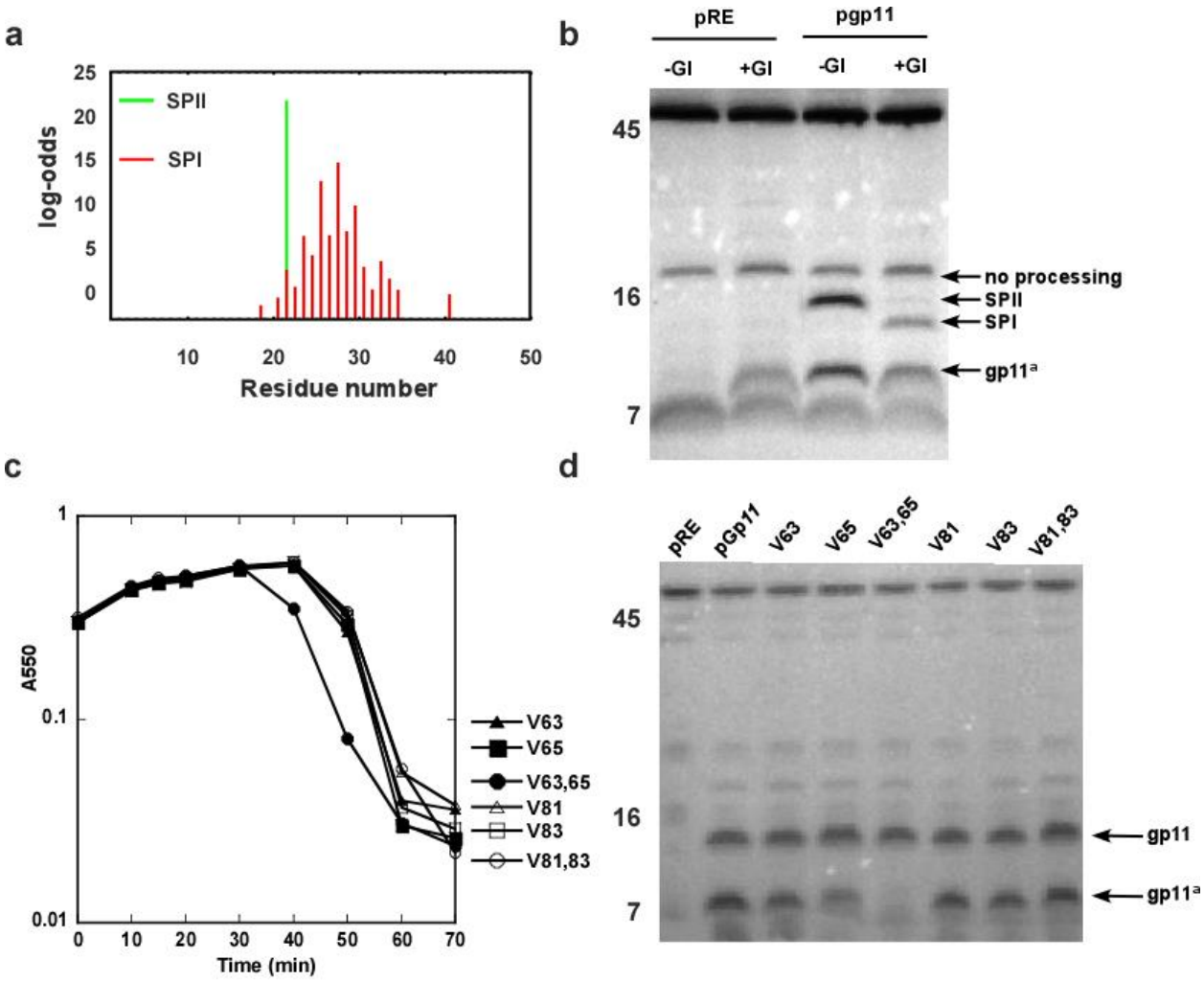
assessing the effect of treatment with globomycin (GI), a specific inhibitor of SPaseII [207], on gp*11* maturation. We could not block complementation of the spanin lysis defect by gp*11* in MC4100 λ$Rz_{am}Rz1_{am}$ cultures even with addition of GI at a concentration as high as 24 μM (Fig. S3.1). Cells treated with higher levels of GI show a loss of the cellular rod shape [208], thus making it difficult for us to differentiate between lysis and effect of GI on the integrity of the

cellular envelope. However, western blotting of whole-cell samples induced for gp*11* expression with IPTG (See methods) indicated that the production of mature gp*11* was severely reduced in the presence of 100 μM GI (Fig. 3.3b). Indeed, it also resulted in the accumulation of a smaller product, thus suggesting gp11 could potentially be a substrate for processing by SPaseI.

A downstream alternative start site

A band of ~7 kDa (denoted by gp*11*[a] Fig. 3.2d) was always present in western blots of membrane samples containing gp*11*. Since the epitope used for immunoblotting is at the extreme C-terminus of gp*11*, we hypothesized that the smaller product was generated by translational initiation downstream of the gene *11* start. There are four potential internal start sites in the gp*11* coding sequence, each a GTG (valine) codon with a reasonable Shine-Dalgarno sequence at an appropriate spacing: codons Val63, Val65, Val81, and Val83 (Fig. 3.2a). To ascertain where the translation of the smaller product started, we created silent mutations eliminating the GTG in each of the Val start codons. None of the individual mutations had any effect on the lytic function of gp*11* or the presence of the smaller species (Fig. 3.3c,d). However, the double mutant allele with synonymous GTG to GTC codon changes at both Val63 and Val65 lacked the smaller species (Fig. 3.3d) and exhibited an altered lytic phenotype: earlier lysis in the complementation assay (Fig. 3.3c). Thus, the low molecular weight species was a result of translational initiations at Val63 and Val65 and the translational products of these downstream initiation events exert a dominant-negative lysis-delay. (See Discussion).

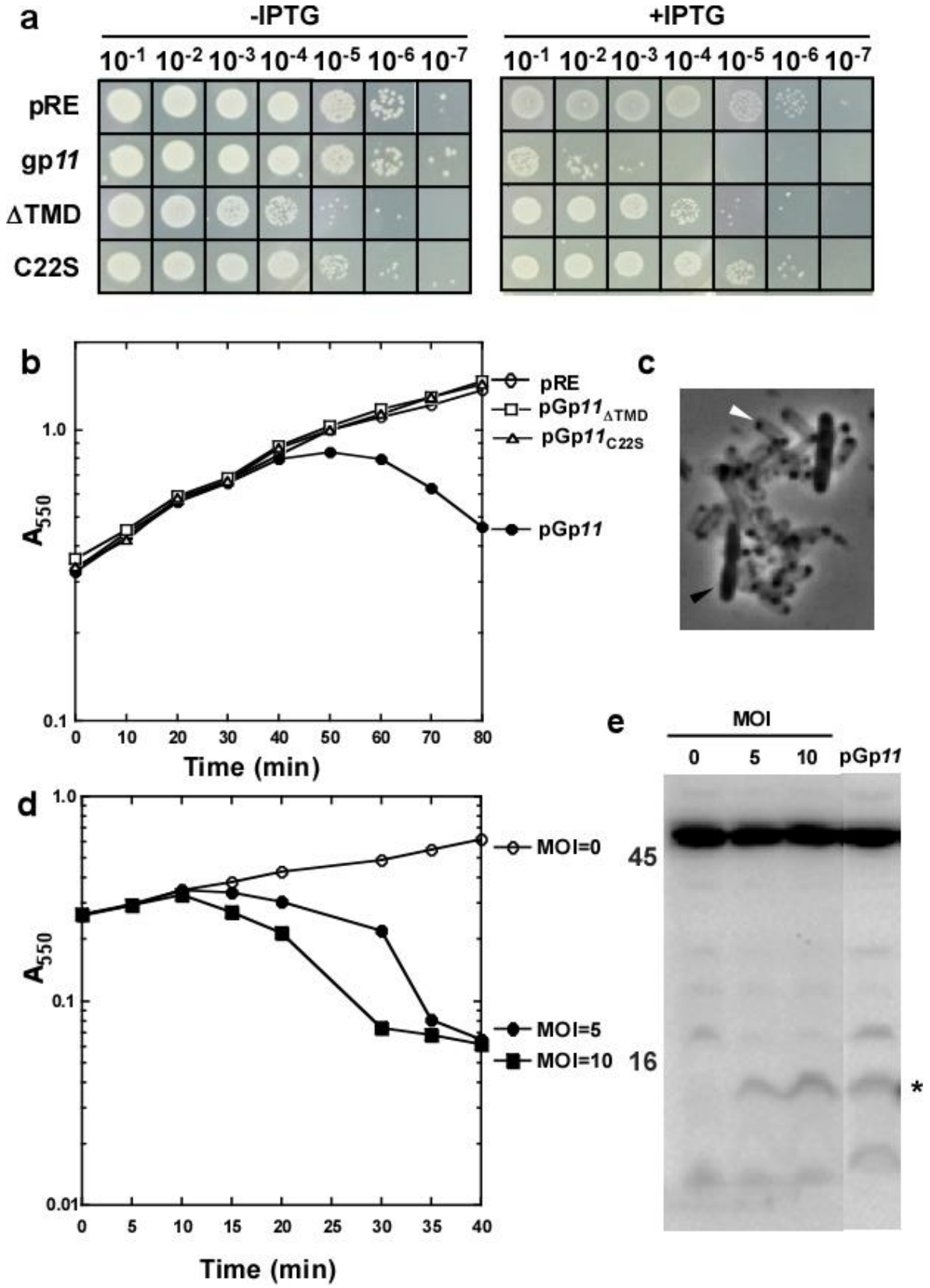
Figure 3.3 Alternative start sites in *gp11*. a) A graph showing the probability of SPaseII (green bars) and SPaseI (red bars) processing at different residue sites of gp*11* as predicted by LipoP 1.0. b) MG1655 *lacI*^{q} pQ cells carrying either pRE or pGp*11* were induced for expression with 1mM IPTG in the presence or absence of 100 μM globomycin. TCA samples were collected at 40 minutes after induction and analyzed by western blotting. The bands corresponding to products cleaved by SpaseII and SPaseI, along with gp*11*^{a} are indicated by arrows. The very faint band corresponding to the uncleaved product is also shown. c) Lysis phenotype of gp*11* internal start site mutants is shown. Cultures of MC4100 $\lambda Rz_{am}Rz1_{am}$ carrying pgp*11* plasmids carrying different combinations of silent mutations at amino acid positions 63, 65, 81 and 83 were induced for lysis as mentioned before and the optical density was followed over time. The silent mutations in each of the constructs are at the following positions: closed triangles = Val63, closed squares = Val65, closed circles = Val 63 and 65, open triangles = Val81, open square = Val83, open circles = Val 81 and Val83. The onset of lysis is for the Val 63,65 mutant starts at 30 min after induction, compared to 40 min for all the other mutants. d) MG1655 *lacI*^{q} pQ cells carrying either pRE or pGp*11* plasmids with above mentioned silent mutations were induced for expression with 1mM IPTG. TCA samples were collected at 40 minutes after induction and analyzed by western blotting. The bands corresponding to the mature gp*11* product and gp*11*^{a}, the smaller molecular weight product due to translation from an alternate start site, are indicated by arrows. The gp*11*^{a} band disappears only when both V63 and V65 start sites are substituted with non-start codons.



Plasmid-borne gp11 expression levels are comparable to T1 infections

In our experiments testing the ability of gp*11* to complement λ$Rz_{am}Rz1_{am}$ lysis defect, we noticed the timing of lysis onset changed from ~50 min in the case of *Rz-Rz1* to 40 minutes in the case of the gp*11* allele used in this study (pGp*11*). We considered that gp*11* might affect the cytoplasmic membrane integrity of the host cell and thus cause earlier lysis. Indeed, expression of gp*11* alone, in the absence of the prophage, reduced viability by ~4 orders of magnitude (Fig. 3.4a). This inducible lethality could be also seen during log-phase growth in liquid cultures, as

Figure 3.4 Comparison of gp*11* expression levels from pGp*11* and T1 infection. a) Toxicity of gp11 expression. Serial dilutions of overnight cultures of MG1655 *lacI*^{q} pQ cells carrying either pRE (open circles) or pGp*11* (WT, ΔTMD, C22S) alleles were made and 5 μl of each diluted was spotted on LB Agar plates with or without IPTG. b) Expression profiles of gp*11* and the topological mutants of gp*11*. MG1655 *lacI*^{q} pQ cells carrying either pRE (open circles) or pGp*11* alleles (closed circles = gp*11*, open square = gp*11*$_{\Delta TMD}$, open triangles = gp*11*$_{C22S}$,) were induced for expression with 1mM IPTG and the optical density was observed over time. c) Micrograph of cells expressing pGp*11* after induction with 1mM IPTG for 60 min. The black and white black arrows indicate an intact cell and a cell that has lost its cytoplasmic contents, presumably due to the fusion of IM and OM by gp*11*. d) Lysis profiles of infection with phage T1. MC4100 cells at an O.D 550 of ~0.25 were infected with phage T1 at a MOI of 5 (closed circles) and 10 (closed squares) and the optical density was observed over time. An uninfected culture (MOI=0: open circles) was used as a control. TCA precipitates of MC4100 cells infected with phage T1 at different MOI were collected at the onset of lysis 40 and analyzed by western blotting to compare it to expression levels from the pGp*11* plasmid. The band corresponding to the mature gp*11* product is indicated by *


a
-IPTG
+IPTG
10-1 10-2 10-3 10-4 10-5 10-6 10-7
pRE
gp11
ΔTMD
C22S
b
pRE
pGp11ΔTMD
pGp11C22S
pGp11
A550
Time (min)
c
d
MOI=0
MOI=5
MOI=10
A550
Time (min)
e
MOI
0 5 10 pGp11
45
16
*

cultures induced for gp*11* expression showed a decrease in optical density (OD) after ~50 min of induction (Fig. 3.4b). The toxic effect was dependent on both the membrane localization signals, as both gp*11*$_{\Delta TMD}$ and gp*11*$_{C22S}$ did not show any decrease in OD in liquid cultures or inducer plates monitored over the same time-period (Fig. 3.4a,b). Phase-contrast microscopy of samples from log phase growth cultures revealed that cells killed by gp*11* induction were transformed into rod-shaped ghosts (Fig. 3.4c), indicating that the cytoplasmic contents had leaked out even though the PG was intact. We wanted to make sure the expression levels of our gp*11* construct were physiologically relevant and comparable to the expression of the spanin in a T1 phage infection. Thus, we followed the lysis kinetics of MC4100 cells infected with T1 phage (Fig. 3.4d) and collected whole-cell samples at the onset of lysis. Western blotting revealed that gp*11* levels at the time of lysis in T1-infected cells were comparable to the level of expression from pGp*11* induced in trans from a λ lysogen (Fig. 3.4e).

Dominance of localization signal mutants

Based on current models for the function of the Rz-Rz1 two-component spanin system, it was considered likely that the u-spanin would also have to oligomerize to effect disruption of the OM (Fig. 3.1c). This suggested that some gp*11* lysis-defective missense alleles should show a dominant phenotype. To test the dominance or recessiveness of the gp*11* mutants, we constructed λ::*11* hybrids (see methods), where the entire *Rz* gene was replaced with one of the non-functional topological mutant alleles: gp*11*$_{\Delta TMD}$, gp*11*$_{C22S}$ or gp*11*$_{IRS}$. Lysogens carrying these hybrid prophages were transformed with pGp*11* and induced for lysis. All the topological mutants exhibited a significant retardation on the lysis mediated by gp*11* expressed in trans (Fig. 3.5a). This partial dominance was not observed when the loci of the alleles was exchanged, i.e.

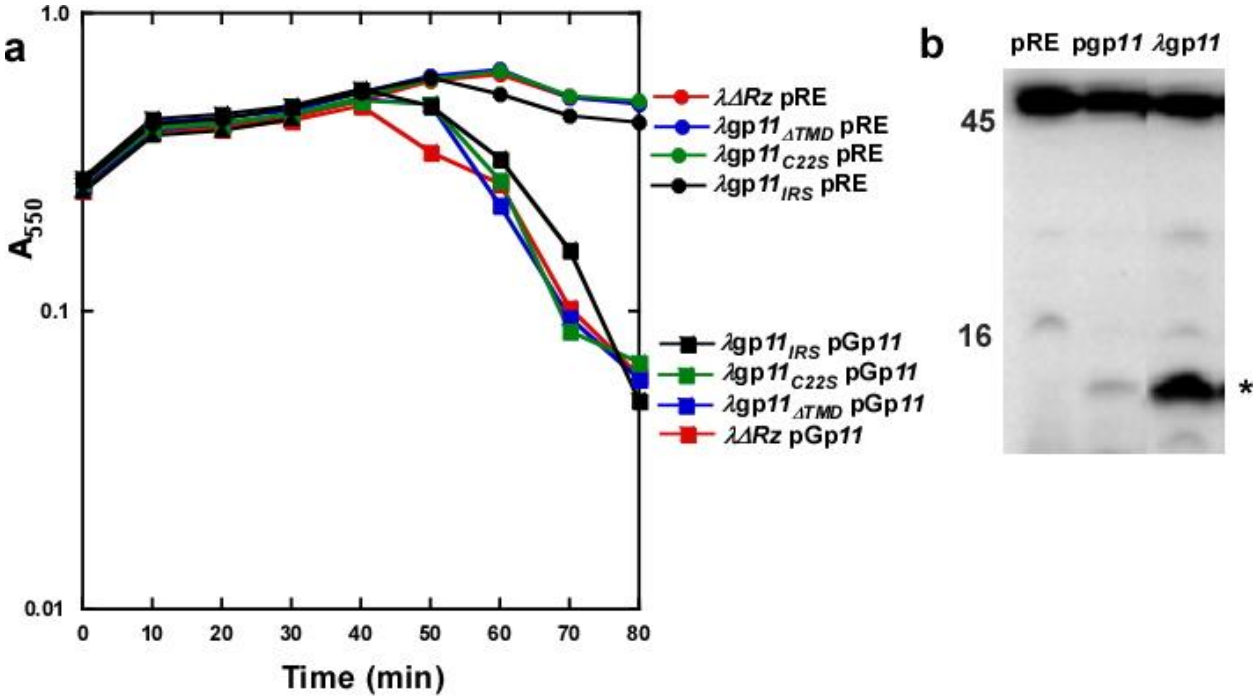


Figure 3.5 Dominance analysis of gp*11* topological mutants. a) Lysis profiles showing dominance of topological mutants. Cells carrying lambda lysogens with no spanin (λΔRz - red) or different topological alleles of gp*11* ($\lambda gp11_{\Delta TMD}$ - blue, $\lambda gp11_{C22S}$ - green and $\lambda gp11_{IRS}$ - black) were transformed with pRE (open circles) or pGp*11* (squares) plasmids and thermally induced for lysis. All three topological alleles had a reproducible delaying effect on lysis by gp*11*. b) Expression levels of gp*11* when expressed from the pGp*11* plasmid and from the lambda chromosomal copy. The band corresponding to the mature gp*11* product is indicated by *

gp*11* expressed from the lysogen and the topological mutants were expressed from the plasmid, implying that the phenotype was dosage dependent. Indeed, western blotting confirmed accumulation of gp*11* expressed from the prophage chromosome was almost 10 times compared to the expression from pGp*11* (Fig. 3.5b).

Visualization of gp11 localization in vivo

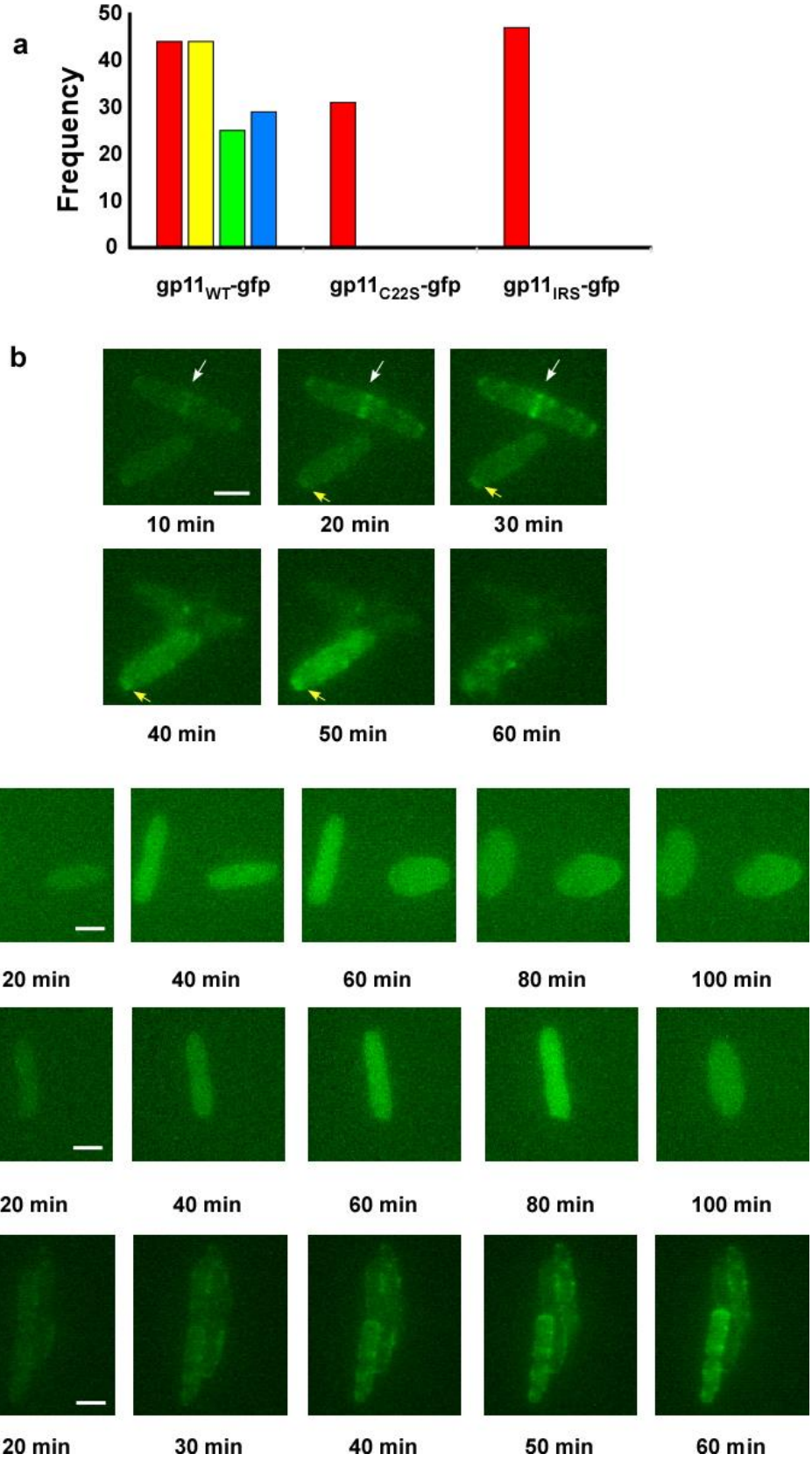
To observe the accumulation of gp11 in the membranes, we created a chimeric allele, gp11-gfp, encoding a product in which the GFP domain is fused to the short, cytoplasmic C-

terminus of gp11. This chimera exhibited full complementation capacity when expressed in trans to an induced $\lambda Rz_{am}Rz1_{am}$ prophage and had a lysis profile almost indistinguishable from the parental allele in a shaker flask (Fig S3.2). Using the GFP fluorescence marker, we noticed that gp*11*accumulated in punctate spots at poles and especially at the mid-cell where, remarkably, it formed rings (Fig. 3.6a,b, Movie S3.1). The punctate distribution is dependent on both the membrane signals, as we found gp11_{C22S}-GFP and gp11_{IRS}-GFP both lost the accumulation phenotype and instead have uniform distribution across the cells (Fig. 3.6a,c,d and Movie S3.2, 3.3). To test whether the punctate distribution of gp*11*-GFP was dependent on the context of the lambda prophage induction, the fusion allele was induced in the non-lysogenic context. Under these conditions, the punctate distribution of gp*11* was still observed but neither preferential accumulation at the mid-cell region nor ring formation was observed (Fig. 3.6e and Movie S3.4). Even in the non-lysogenic context, the punctate distribution is dependent on both the membrane signals, as we found that gp11_{C22S}-GFP and gp11_{IRS}-GFP both showed uniform distribution throughout the cells (Movie S3.5 and S3.6). Prolonged induction resulted in the production of ghosts, where the cytoplasmic contents had been leaked out.

The PG meshwork acts as a negative inhibitor of gp11 function

Previously, we showed that PG depletion in the presence of physiological levels of Rz/Rz1 resulted in lysis, independent of holin or endolysin function [117]. This implied the intact PG network acts a negative regulator of two-component spanin function, preventing IM-OM fusion until the lytic pathway is initiated by holin triggering. To test whether the u-spanin was subject to the same regulation, we repeated this PG depletion experiment with gp*11* as

Figure 3.6 Accumulation of gp*11*-GFP. a) Histogram showing the distribution of gp*11*-GFP foci. Cultures of MC4100 $\lambda Rz_{am}Rz1_{am}$ carrying pGp*11-gfp* or the topological mutant versions, gp11_{C22S}-*gfp* and gp11_{IRS}-*gfp*, were thermally induced for lysis in shaker flask cultures. After 15 min, cells were placed on an agarose pad containing $MgCl_2$ and the accumulation of gp*11*-GFP was followed by fluorescence microscopy. Number of cells showing GFP fluorescence (red bars), punctate distribution (yellow bars), mid cell accumulation (green bars) and polar accumulation (blue bars) were counted for each population and plotted. b) Representative snap shots of accumulation of gp*11*-GFP in thermally induced MC4100 $\lambda Rz_{am}Rz1_{am}$ cells over time. White and yellow arrows indicate the mid-cell and polar accumulation of gp*11*-GFP. White bar in the first snapshot represents a scale bar, 2 μm. c,d) Representative snap shots of accumulation of gp11_{C22S}-GFP (c) and gp11_{IRS}-GFP (d) in thermally induced MC4100 $\lambda Rz_{am}Rz1_{am}$ cells respectively over time. Due to the pressure of being under the agarose pad, cells appear oval instead of the spherical phenotype characteristic of spanin null mutants. White bar in the first snapshot represents a scale bar, 2 μm.


a
Frequency
50
40
30
20
10
0
gp11WT-gfp
gp11C22S-gfp
gp11IRS-gfp
b
10 min
20 min
30 min
40 min
50 min
60 min
c
20 min
40 min
60 min
80 min
100 min
d
20 min
40 min
60 min
80 min
100 min
e
20 min
30 min
40 min
50 min
60 min

reported by Rajaure et al., we used a *dadX alr* host carrying a $\lambda S_{am}RRz_{am}Rz1_{am}$ prophage which allows us to test gp*11* function by depleting the PG in an endolysin-independent manner. The gene products of *dadX* and *alr* are alanine racemases essential for PG biosynthesis, responsible for conversion of L-Ala to D-Ala [209]. As with the *RzRz1* system, induced cultures of cells

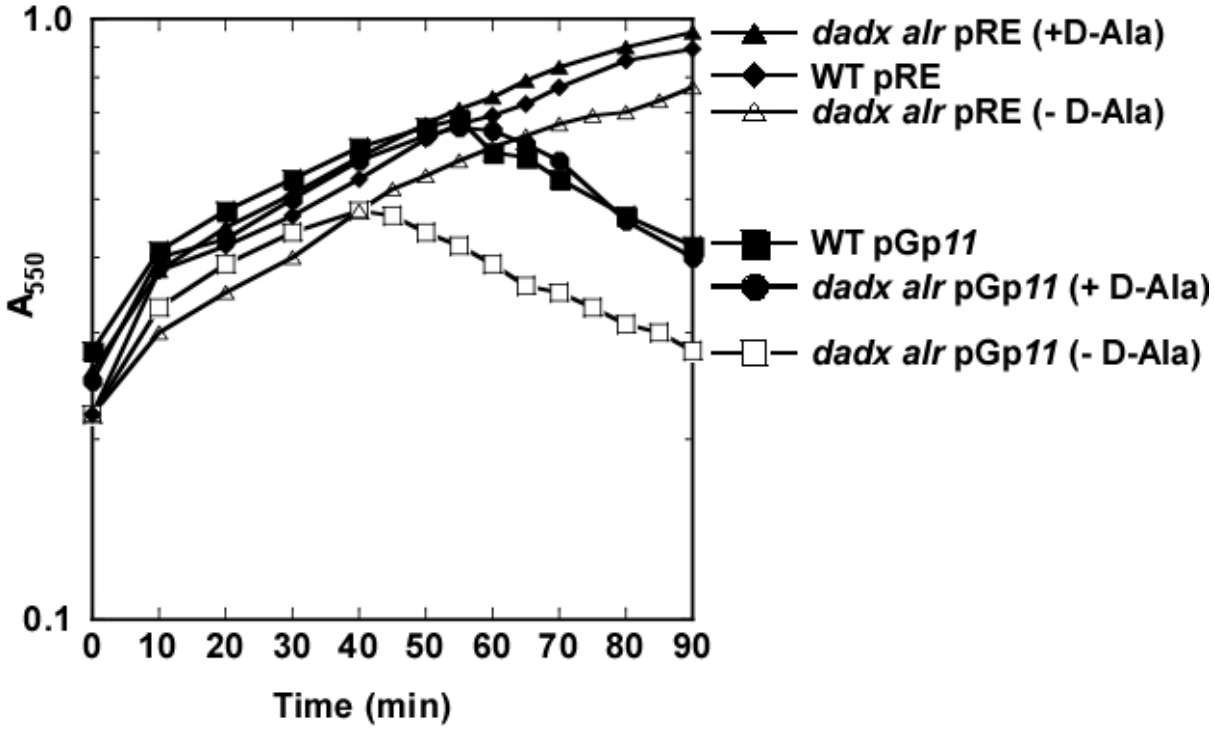


Figure 3.7 Spanin function is negatively regulated by the PG. (A) Cells carrying the $\lambda S_{am}RRz_{am}Rz1_{am}$ prophage and auxotrophic for D-Ala were transformed with pRE or pGp*11* and grown in LB supplemented with D-Ala until an OD of ~0.2 and thermally induced for lysis. An isogenic host which is not auxotrophic for D-ala was used as a wild type control. To observe the effect of D-Ala depletion, cells were collected by centrifugation and washed with fresh LB three times and resuspended in LB without D-Ala to the same volume before thermal induction. While the cultures expressing pRE (*dadXalr* pRE without D-Ala: open triangles, *dadXalr* pRE with added D-Ala: closed triangles, WT pRE: closed diamonds) continued to grow irrespective of the host type, cells expressing pGp*11* showed a decrease in O.D over time. The wild type host (closed squares) and the auxotroph supplemented with D-Ala (closed circles) both followed similar growth kinetics, showing a decrease of O.D around 55 min. However, the auxotroph without any supplemental D-Ala showed a sharp change in the onset of lysis at about 40 min.

carrying the spanin gene in trans to the lysis-defective prophage pGp*11* in the absence of D-Ala resulted in a sharply-defined lysis event as early as 40 min (Fig. 3.7). In comparison, both an isogenic control prototrophic for D-Ala and the auxotroph supplemented with D-Ala showed lysis beginning at ~55 min. Thus, PG depletion can activate the gp*11* u-spanin to cause lysis without holin-endolysin function, in support of the membrane fusion model.

Discussion

Here, we report the first functional characterization of a u-spanin, the prototype gp*11* from phage T1. The key findings must be considered in the context of the current understanding of the biosynthetic and topological pathways of the Gram-negative envelope.

gp11 has functional IM and OM localization signals, loss of either leads to dosage dependent-dominant lysis-defective phenotypes

These findings raise questions regarding the processing of gp*11*. Of the ~1000 membrane proteins encoded in *E. coli*, only 17 are predicted to have a single C-terminal TMD with the Type I N-out, C-in topology like Gp*11* (Table S3.1). Five of these have small N-terminal domains that would not require the *sec* system for export [210, 211]. Of the remaining 11, only three have legitimate signal peptidase I or signal peptidase II (SPII) signal sequences, suggesting the other 8 are erroneously predicted to a have a single TMD or, in these cases, the TMD is actually in N-in, C-out topology and thus belongs to the class of tail-associated membrane proteins that are imbedded in the IM by a *sec*-independent, YidC-dependent process [212, 213]. Only one protein, YhdV (NP_417733.1) has a legitimate lipobox and a predicted SPII cleavage site; even here, there is a predicted SPI cleavage site, so the localization is ambiguous. If SPII-

processed, YhdV is predicted to be an OM lipoprotein, based on its Ala residue immediately following the presumptive Cys lipoylation residue. However, the mature length of processed YhdV would only be 59 residues, with only 36 N-proximal to the predicted C-terminal TMD; using 0.38 nm per residue for a maximally extended peptide, the periplasmic domain of YhdV would only extend ~14 nm, which would be insufficient to span the periplasm. This indicates that YhdV localization would depend on a competition between OM and IM localization of the processed lipoprotein form; more likely, YhdV may be primarily processed by SPI, resulting in a C-terminally tethered Type I IM protein. Thus, in a T1-infected *E. coli* cell, Gp*11* processing would be unique, in that two different localization pathways, the OM lipoprotein maturation and OM export pathway and the *sec*-dependent integration of the TMD into the IM, both occur, either simultaneously or in a kinetically-determined sequence. (See below.)

Oligomerization in gp11 function

Two lines of evidence indicate that gp*11* oligomerizes as part of its lytic function. First, defects in both IM and OM localization signals lead to dosage-dependent dominant-negative phenotypes. Second, gp*11*-GFP fusions, which show normal function in complementation tests, accumulate in discrete foci under physiological levels of expression (Fig. 3.6b). This was expected, since there is compelling evidence that the two-component spanins undergo oligomerization [52, 129], but it does raise questions about how gp*11* is localized. If the transfer to the OM occurs after integration of the gp*11* TMD, then "inchworm" forms of gp*11*, with both lipoylated N-terminus and C-terminal TMD inserted in the IM, may accumulate and multimerize via two-dimensional diffusion in the IM. This notion is especially attractive since specific interactions that could support homo-oligomerization would have to be in the periplasmic

domain, in view of the ability of the synthetic TMD to replace the parental TMD. Once the oligomers are formed, the Lol system could export the lipoylated N-termini to the OM Gp*11* one molecule at a time, allowing punctate foci of periplasm-spanning gp*1l* molecules to form. In contrast, if Gp*11* is exported through the PG to the OM before the TMD escapes laterally into the IM, each molecule would presumably be trapped in a single lacuna of the PG network, which would make oligomerization into foci conceptually difficult. Homo-oligomerization has already been demonstrated for the Rz-Rz1 two-component spanin system, where the presence of coiled-coil domains provides a ready mode for protein-protein interactions. In contrast, the gp*11* periplasmic domain is dominated by predicted beta-strand segments, implying the interactions between the subunits might involve mixed strand beta-sheet structures.

Analysis of the periplasmic domain

Confirmation of the requirements for integration into both membranes indicates that the 85-residue segment between Cys22 and the C-terminal TMD crosses the periplasm. Secondary structure predictions indicate 14% of the residues in the periplasmic domain adopt an alpha helical conformation while 31% of them arrange themselves into beta strands (Fig. 3.2a). Assuming the residues in an alpha helix contribute to a distance of 0.15 nm each, the predicted alpha helix would cover ~2 nm, leaving 73 residues to account for the rest of the distance between the membranes. Given the ~25 nm width of the periplasm [49, 52], it is conceivable that the rest of gp*11* periplasmic domain would need to be in a fully extended confirmation, spanning ~0.38 nm per residue, to account for the remaining ~23 nm. The oligomerization of gp11 molecules, followed by the conformational changes upon degradation of the PG, could result in

condensation into beta sheets or beta barrels, shortening the length dramatically, and thus facilitating membrane fusion.

BLAST analysis of gp*11* revealed only predicted u-spanin homologs from other phages. Since no insights could be gained from sequence similarity searches, we looked for proteins with secondary structure patterns similar to that predicted for gp*11*. One such protein is YuaF, a member of the NfeD-like (Nodulation formation efficiency D) class of proteins [214]. The C-terminal fragment of YuaF, residues 97-174, is predicted to have an alpha helix followed by five beta sheets; this secondary structure was confirmed using high resolution NMR [215]. According to Walker et.al, this soluble fragment of YuaF (sYuaF) adopts an oligonucleotide/ oligosaccharide-binding-protein (OB) fold topology, wherein the five beta sheets arrange themselves into a closed beta barrel structure. While gp*11* and sYuaF share no significant sequence similarity, it is possible the periplasmic domain of gp*11* assumes a similar topology as that of sYuaF, given the secondary structure pattern. Change from an initial stretched/extended conformation to a collapsed conformation, such as the OB fold, would serve the purpose of bringing the IM and OM together, eventually resulting in membrane fusion. OB-folds are known to frequently oligomerize or are found in large multicomponent assemblies [216], which makes the idea of gp*11* adopting such a conformation plausible. With the potential of gp*11* adopting the OB-fold, it can also be hypothesized that interaction with the PG fragments resulting from endolysin activity could act as a driving force for the conformational change required for IM-OM fusion.

Punctate localization of gp11-GFP indicates clustered assembly of the u-spanin

The N-terminal lipoylated part of gp*11* depends on the host Lol system for transport to the OM. Once the periplasmic transporter LolA transfers the lipoylated cysteine end from the IM to OM, mature gp*11* molecules are trapped in the lacuna they are threaded through. Periplasmic interactions between the beta sheet structures of adjacent gp*11* molecules within the same mesh-hole of the PG network could lead to gp*11* oligomers and thus the punctate GFP foci. However, it is not clear yet, why the punctate distribution is not spread throughout the membranes. This could be attributed to localized PG turnover or the localized concentrations of players involved in the Lol export system, especially the LolCDE complexes in the IM. The oligomerization does depend on the extended conformation of gp*11* as gp*11*$_{IRS}$ did not show any punctate foci. The other interesting finding was the mid-cell ring accumulation phenotype when gp*11* was expressed in a λ prophage background. It can be interpreted as an artefact of the *kil* gene expression in the λ lysogen, which inhibits FtsZ ring formation and results in filamentous cells [217]. Assuming Kil function indirectly affects PG remodeling/turnover at mid-cell, PG lacunae in this region could be wider than other regions of the cell, allowing more gp*11* molecules to accumulate within each lacuna, thus giving us the mid-cell ring phenotype. While we cannot rule out the possibility of the mid cell accumulation of gp*11* as physiologically relevant for lysis, cells expressing just gp*11* show no preferential mid-cell accumulation and can still achieve the proposed fusion of IM and OM. Irrespective of the context, the punctate foci indicate local oligomeric concentrations might prime the u-spanins for rapid OM disruption once the PG is removed.

gp11, like the Rz-Rz1 complex, can effect lysis without holin-endolysin function if PG is depleted

The findings with the D-Ala auxotroph indicate u-spanins can effect "holin-endolysin"-independent lysis upon partial PG degradation. This suggests that the step of OM disruption by gp*11* is not coupled to the preceding lytic steps dealing with the IM and the PG. Leakage of cell contents upon expression of gp*11* alone also supports this notion. These results are in accordance with our observations of Rz-Rz1, the two-component spanins from lambda [117]. It is important to note that, in these situations, the cells retain their rod-shaped morphology. The simplest interpretation is that both types of spanins use the topological solution of IM-OM fusion to subvert the OM. This also supports the notion of PG acting as an inhibitor of spanin function until the ideal time for host lysis, which is dictated by the release of endolysin by the holin. From this perspective, the importance of regulating gp*11* function is paramount, as premature membrane fusion would terminate the phage infection cycle. Thus, the physical trapping of spanin complexes by PG could be considered a damping mechanism against premature IM-OM fusion. Even though intrinsically different from the λ two-component spanins Rz-Rz1, in terms of secondary structure gp*11* seems to achieve the same end result and to be regulated in a similar fashion by the PG. The parallels also extend to eukaryotic viral fusion systems [153]. The two-component systems can be considered analogous to the class I viral membrane fusion proteins, which are dependent on alpha helical coiled-coils for their function. With the predicted beta sheet secondary structures, gp*11* and other u-spanins would be comparable to class II viral membrane fusion proteins, which consist of globular domains made up of beta sheets. Similar to the oligomerization and series of conformational changes in the SFV E1 ectodomain induced by low pH [155], the trigger in the case of u-spanins could be the breakdown of PG.

Not only does gp*11* seem to be capable of effecting lysis by itself, but also alter the timing of lysis in the lambda prophage context. An optimal lysis time is vital for phage propagation and changes can affect burst size and phage fitness [90, 92]. Holins accumulate in the IM harmlessly until reaching a threshold concentration, after which "triggering" occurs, forming lesions in the IM [92]. This timing of triggering is encoded into the primary structure of holin and single missense mutations in the holin sequence result in a varied spectrum of lysis times, ranging from very early to very late/blocked lysis [91, 105, 218]. Moreover, lysis proteins are thought to function independently of each other; i.e. the presence or absence of endolysin and spanins does not affect holin triggering time. However, triggering can be induced prematurely by the addition of energy poisons, like potassium cyanide, which collapse the proton-motive force across the IM [104, 106]. Thus, the simplest interpretation is that accumulation of gp11 has an impact on the energization and membrane integrity of the IM, thus indirectly influencing holin triggering.

Downstream translational starts generate "anti-spanin" gene products

The product of an internal translational start gp*11*^{a} (Fig. 3.3) has a negative dominant effect on gp*11* function, suggesting the shorter and mislocalized gene product serves as an intrinsic "anti-spanin", providing another level of regulation for u-spanin function. This is reminiscent of the "dual-start" holin genes, lambda *S* and phi21 *S21*, in which separate translational starts are used to generate anti-holin products [110, 121, 219]. However, in these cases the anti-holins are expressed from upstream translational starts, in contrast to the downstream start as in the case of gp*11*. Inhibition of the SPaseII processing using GI revealed that gp*11* could act as a substrate for SPaseI processing, resulting in a non-lipoylated product.

Given the negative dominant effect of the topological mutants of gp*11*, such as the gp*11*C22S, it is possible the mature product of gp*11* cleaved by SPaseI is physiologically relevant. The competitive processing of gp*11* between the SPases could thus act as another step of negative regulation of u-spanin function. Thus, there may be three different levels of regulation of u-spanin function: the physical barrier provided by the PG; the negative regulation from the non-lipoylated gp*11* produced by SPaseI cleavage; and the downstream start anti-spanin gp*11*^{a} product. The smaller u-spanin products potentially interact with the extended periplasmic domain of full length gp*11*, essentially poisoning the functional gp*11* oligomers. This perspective suggests that trans inhibitors, based on the periplasmic domain of gp*11*, could be designed to block the disruption of OM entirely.

Materials and methods

Bacterial strains, phages, plasmids, and growth conditions

The bacterial strains, bacteriophages, and plasmids used in this study are described in Table 3.1. Bacterial cultures were grown in standard LB medium, supplemented with antibiotics; ampicillin (Amp, 100 μg/ml) and kanamycin (Kan, 40 100 μg/ml) as appropriate. Expression of the pRE plasmid is governed by the lambda late promoter pR'. Thus, to activate pR', the antiterminator Q is supplied in trans by either the induced prophage or the p*Q* plasmid. For experiments to test complementation of the *λRzRz1* lysis defect, MC4100 (*λ*900$Rz_{am}Rz1_{am}$) cells were transformed with the pRE-plasmid carrying the concerned allele and transformants selected on Amp-Kan plates at 30°C were used to inoculate overnight cultures. The overnight culture was diluted 1:250 into 25 ml fresh LB media and grown to an OD of ~0.25 at 30°C. Lysis was thermally induced by a shift from 30°C to 42°C for 15 min, followed by continued growth at 37°C. $MgCl_2$ was added to a final concentration of 10 mM at the time of induction to stabilize the outer membrane. Growth and lysis of cultures over time was monitored by A_{550} using the Gilford Stasar III spectrophotometer. For expression of gp*11* in non-lysogenic context, MG1655 *ΔtonA* $lacI^{q1}$ p*Q* cells were transformed with the pRE-plasmid carrying the concerned allele and transformants selected on Amp-Kan plates at 37°C were used to inoculate overnight cultures. The overnight culture was diluted 1:250 into 25 ml fresh LB media and grown to an OD of ~0.2 before inducing by addition of isopropyl β-D-thiogalactopyranoside (IPTG) was added to the final concentration of 1 mM for plasmid induction. For the globomycin experiments, GI was added at mentioned concentrations to cultures at the point of induction of gp*11* expression. Lysis and growth curves were plotted using Kaleidagraph.

DNA manipulation and construction of λ::11 hybrids

The primers for site directed mutagenesis used in this study are listed in Table 3.2. All primers were ordered from Integrated DNA technologies (Coralville, IA). Plasmid isolation was performed using the Qiagen spin miniprep kit and mutations were confirmed by sequencing results from Eton Biosciences (San Diego, CA).

Construction of lambda hybrids with modified lysis cassettes was performed as described before [128]. First pRK1 was constructed from pER157 (which carries the lambda lysis cassette followed by *Δbor::kan* insertion) by deleting the *Rz* gene via PCR slicing using the λ *ΔRz* primers. The *gp11* gene was amplified using the λ *ΔRz::gp11* primers on the pGp*11* plasmid, and the amplicon was purified using a Qiagen PCR purification kit. The purified amplicon was then used as a "mega-primer" to insert the *gp11* gene into pRK1, resulting in pRK2. These plasmids were then used to recombine the modified lambda lysis cassette with either no spanin or the u-spanin gp*11* into the prophage chromosome. Briefly, the plasmids carrying the modified lambda lysis cassette were transformed into the strain MC4100 (*λΔSR*) and induced for lysis. The phage lysate was used to lysogenize MC4100 and lysogens that were Kan resistant and Amp sensitive were tested for S^+R^+. Single lysogens from the above mentioned positives were identified by screening lysogens using colony PCR [220] and the changes in the lysis cassette were confirmed by sequencing. The pRK2 plasmid was used as a template to create lysis cassette versions carrying topological mutant alleles of gp*11* and then recombined to constructs *λ::11* hybrids carrying the respective signal sequence mutations.

SDS-PAGE and western blotting

SDS-PAGE and western blotting were performed as described previously [105]. TCA pellets of whole cell samples were acetone washed, resuspended in 1X SDS-PAGE buffer with β-mercaptoethanol (100mM BME). Protein samples were loaded onto 10% resolving Tris-tricine polyacrylamide gels after loading volumes were normalized according to A_{550} at the time of collection of TCA precipitates. Proteins were transferred to PVDF membrane (Pall Life Sciences) using a Hoefer TE unit at 0.5 mA for 2 hours. Antibodies (Genscript) were generated in rabbits against a synthetic peptide CIRKHEKKEA, corresponding to the C-terminal end of gp*11*. The primary antibodies were used at a dilution of 1:1000 while the secondary antibody, goat-anti-rabbit-HRP (Thermo Scientific), was used at a dilution of 1:5000. Chemiluminescence was detected using a Bio-Rad XR Gel Doc system. SeeBlue Plus2 (Invitrogen) pre-stained standard served as a molecular mass standard.

Fluorescence microscopy

MC4100 (*λ*900*Rz*$_{am}$*Rz1*$_{am}$*)* cells carrying the different topological mutant alleles of pGp*11-gfp* were grown to an OD of ~0.25 at 30°C. Lysis was thermally induced by a shift to 42°C for 15 min, and then 1 μl of sample was placed onto a coverslip (24 × 50 mm; thickness, 0.16 mm). After gently covering the sample with an agarose pad (1% wt/vol in LB+ 10 mM $MgCl_2$) of ~5 × 5 mm dimensions, the cells were continued to grow at 37°C while imaging was performed simultaneously on an inverted microscope (Ti-E, Nikon, Tokyo, Japan) with a cage incubator (InVivo Scientific, St. Louis, MO) set at 37°C. Images were taken using 100× objective (Plan Fluo, NA 1.40, oil immersion) with standard filter sets and a cooled EMCCD camera (iXon 3

897, Andor, Belfast, United Kingdom). Images were taken every 1 min for the specified times through the phase-contrast and GFP channels (100 ms exposure) over 8-10 different coordinates. Image analysis and processing was done using NIS-Elements imaging software. For the mid-cell, polar and punctate accumulation statistics, each image from different coordinates was analyzed and cells exhibiting the desired phenotype were manually counted and plotted using Kaleidagraph. The final movies showing the phase, gfp and merged channels together were created using Windows Movie Maker. For visualizing the accumulation of gp11 in a non-lysogenic background, MG1655 *lacI*^{q} pQ cells carrying pGp*11*-gfp alleles were grown at 37°C to an OD of ~0.25 and induced with 1mM IPTG. After 15 min of induction, 1 μl of culture was placed under an agarose pad with IPTG (1% wt/vol in LB+ 1 mM IPTG) and visualized as described above.

PG depletion experiments

The D-Ala depletion experiments were performed as described before in [117]. Briefly, MC1000 (*λ*901*Rz*$_{am}$*Rz1*$_{am}$) and MB2159 (*λ*901*Rz*$_{am}$*Rz1*$_{am}$) were used as the isogenic wild type and racemase mutant (*dadX alr*) hosts. Competent cells of the above strains were transformed with pRE or pGp*11* and selected on Amp-Kan plates supplemented with 150 μM D-Ala. Overnight cultures of these transformants were diluted 1:300 into fresh LB media and 150 μM D-Ala and grown at 30°C until an O.D of ~ 0.25 before thermal induction. For depleting D-Ala condition, the culture was harvested by centrifugation at 6,500 × g for 5 min at room temperature using a Thermo Scientific F15S-8 × 50cy rotor, washed three times with LB to remove the remaining D-Ala in the media, and resuspended in fresh media before thermal induction.

Tables

Table 3.1 Phages, bacterial strains and plasmids

Name	Description	Reference
Phages		
λ900	*λΔ(stf tfa)::cat cI857 bor::kan*; carries Cam^R and Kan^R	Laboratory stock
λ900Rz_{am} $Rz1_{am}$	λ900 carrying the Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles of the spanin genes	[115]
λ901	λ900 carrying the S_{am7} nonsense allele of the S holin gene	Laboratory stock
λ901Rz_{am} $Rz1_{am}$	λ901 carrying the Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles of the spanin genes	[115]
λ *ΔSR*	*stf::cat::tfa cI857 Δ(SR);* carries Cam^R	[128]
λ900 *ΔRz*	λ900 carrying a deletion of the *Rz* gene	This study
λ900 *Rz::gp11*	λ900 with the *Rz* gene deleted and substituted with *gp11*	This study
***E. coli* strains**		
MC4100 *tonA::Tn10*	Escherichia coli K-12 *F araD139 Δ(argF-lac)U169 rpsL15 relA1 flbB3501 deo pstF25 rbsR tonA*	Laboratory stock
MC4100 (*λ*900$Rz_{am}Rz1_{am}$)	MC4100 *tonA::Tn10* carrying λ900 prophage with Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles	[115]
MC4100 (*λΔSR*)	MC4100 *tonA::Tn10* carrying the *λΔSR* prophage	[128]
MC4100 (*λ*900*ΔRz*)	MC4100 *tonA::Tn10* lysogen carrying λ900 prophage with *Rz* gene deleted	This study
MC4100 (*λ*900 *Rz::gp11*)	MC4100 *tonA::Tn10* lysogen carrying λ900 prophage with *gp11* substituted for the *Rz* gene	This study
MC1000	*araD139 Δ(ara-leu)7679 galU galK Δ lac*	[209]
MB2159	MC1000 $dadX_{EC}$*::frt* alr_{EC}*::frt*	[209]
MC1000 (*λ*901)	MC1000 lysogen carrying λ901prophage	[117]
MC1000 (*λ*901$Rz_{am}Rz1_{am}$)	MC1000 lysogen carrying λ901prophage with Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles	[117]
MB2159 (*λ*901)	MB2159 lysogen carrying λ901prophage	[117]
MB2159 (*λ*901$Rz_{am}Rz1_{am}$)	MB2159 lysogen carrying λ901prophage with Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles	[117]
MG1655 *ΔtonA* $lacI^{q1}$ p*Q*	MG1655 *ΔtonA* $lacI^{q1}$ carrying the p*Q* plasmid	[221]

Table 3.1 (continued)

Name	Description	Reference
Plasmids		
pRE	Medium copy vector with the λ late promoter pR' that is transcriptionally activated by λQ; Amp^R	[221]
p*RzRz1*	pRE plasmid carrying the embedded *Rz Rz1* genes	[115]
pGp*11*	pRE plasmid carrying the *T1gp11* gene	[116]
pGp$11_{\Delta TMD}$	pRE plasmid carrying the *T1gp11* allele in which the C-terminal TMD from residues 108 to 124 is deleted	This study
pGp11_{C22S}	pRE plasmid carrying the *T1gp11* allele in which the lipoylation Cys22 residue is substituted with Ser	This study
pGp11_{IRS}	pRE plasmid carrying the *T1gp11* allele in which the Ser23 and Thr24 residues are substituted with Asp	This study
pGp$11_{art\text{-}TMD}$	pRE plasmid carrying the *T1gp11* allele in which the C-terminal TMD from residues 108 to 124 is replaced with the residues VLLIIVVVVVVVVIILLI	This study
p*Gp11-gfp*	pRE plasmid carrying the *T1gp11* allele with a translational GFP fusion at the C-terminus	This study
pQ	λ *Q* gene cloned under $P_{lac/ara\text{-}1}$ promoter in a low copy number plasmid pZS-24; Kan^R	[106]
pER157	pBR322 Δ*tet* (*SRRzRz1 bor*::*kan*); The insert is flanked by regions with homology to the phage chromosome	[128]
pRK1	pBR322 Δ*tet* (*SRΔRz bor*::*kan*)	This study
pRK2	pBR322 Δ*tet* (*SRgp11 bor*::*kan*)	This study

Table 3.2 List of primers

Primer name	Sequence
gp*11*$_{\Delta TMD}$ For	5'- CAAATCTGGAAGCTGGTACGATCCTATTCGTAAGCACGAAAAAAAGGAG -3'
gp*11*$_{\Delta TMD}$ Rev	5'- CTCCTTTTTTTCGTGCTTACGAATAGGATCGTACCAGCTTCCAGATTTG -3'
gp*11*$_{pelB\text{-}ss}$ For	5'- CTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCCTGTTCGACGGCATCTGCTCTGAGTG -3'
gp*11*$_{pelB\text{-}ss}$ Rev	5'- ACCAGCAGCAGCGGTCGGCCGCAGGTATTTCATTATAGTGCCGTAACGCAATATGCG -3'
gp*11*$_{C22S}$ For	5'- CTCTATCCGGTTCTTCGACGGCATC -3'
gp*11*$_{C22S}$ Rev	5'- GATGCCGTCGAAGAACCGGATAGAG -3'
gp*11*$_{IRS}$ For	5'- GTGATTTCTCTATCCGGTTGTGACGACGCATCTGCTCTGAGTGGTTTAC -3'
gp*11*$_{IRS}$ Rev	5'- GTAAACCACTCAGAGCAGATGCGTCGTCACAACCGGATAGAGAAATCAC -3'
gp*11*$_{art\text{-}TMD}$ For	5'- GTGGTTGTGGTGATCATCCTGCTGATTCGTAAGCACGAAAAAAAGGAGGC -3'
gp*11*$_{art\text{-}TMD}$ Rev	5'- CACAACAACAACGATGATCAGCAGCACAGGATCGTACCAGCTTCC -3'
gp*11*$_{V63syn}$ For	5'- CAAAGGCGGATGATAAGCGAGAAGTCAAGGTGAGTGATTCAAATATTG -3'
gp*11*$_{V63syn}$ Rev	5'- CAATATTTGAATCACTCACCTTGACTTCTCGCTTATCATCCGCCTTTG -3'
gp*11*$_{V65syn}$ For	5'- GATAAGCGAGAAGTGAAGGTCAGTGATTCAAATATTGGC -3'
gp*11*$_{V65syn}$ Rev	5'- GCCAATATTTGAATCACTGACCTTCACTTCTCGCTTATC -3'
gp*11*$_{V63,65syn}$ For	5'- CAAAGGCGGATGATAAGCGAGAAGTCAAGGTCAGTGATTCAAATATTGGC -3'
gp*11*$_{V63,65syn}$ Rev	5'-GCCAATATTTGAATCACTGACCTTGACTTCTCGCTTATCATCCGCCTTTG -3'
gp*11*$_{V81syn}$ For	5'- CATCCGTCAAGAAGTCCGTAGAGGTGTCAACCATTCAG -3'
gp*11*$_{V81syn}$ Rev	5'- CTGAATGGTTGACACCTCTACGGACTTCTTGACGGATG -3'
gp*11*$_{V83syn}$ For	5'- GTCAAGAAGTCCGTGGAGGTATCAACCATTCAGGCCAAC -3'
gp*11*$_{V83syn}$ Rev	5'- GTTGGCCTGAATGGTTGATACCTCCACGGACTTCTTGAC -3'
gp*11*$_{V81,83syn}$ For	5'- CTCATCCGTCAAGAAGTCCGTAGAGGTATCAACCATTCAGGCCAAC -3'
gp*11*$_{V81,83syn}$ Rev	5'- GTTGGCCTGAATGGTTGATACCTCTACGGACTTCTTGACGGATGAG -3'
gp*11*-gfp For	5'- CGTAAGCACGAAAAAAAGGAGGCGATGAAACAGCATGACTTTTTCAAG -3'
gp*11*-gfp Rev	5'- GGTACAAATAAGAAAGGAGGCTTATTTGTATAGTTCATCCATGCCATGTG -3'

Table 3.2 (continued)

Primer name	Sequence
λ Δ*Rz* For	5'- GCGGAACGGTCAGAGAGATTGATGTATGAAGTTGCCCATATCGATG GGCAACTCATGC -3'
λ Δ*Rz* Rev	5'- GCATGAGTTGCCCATCGATATGGGCAACTTCATACATCAATCTCTCT GACCGTTCCGC -3'
λ Δ*Rz::gp11* For	5'- GCGGAACGGTCAGAGAGATTGATGTATGAAACTTAAGAAAACGTGC ATTGC -3'
λ Δ*Rz::gp11* Rev	5'- GCATGAGTTGCCCATCGATATGGGCAACTTTACGCCTCCTTTTTTTC GTGCTTACG -3'

CHAPTER IV

GENETIC AND BIOCHEMICAL ANALYSIS OF THE PARADIGM UNIMOLECULAR SPANIN GP11 FROM PHAGE T1

Introduction

The lysis mechanisms of double-stranded DNA phages of Gram-negative hosts have been studied extensively for decades [101, 102]. After the phage holin and the endolysin function to overcome the inner membrane (IM) barrier and destroy the peptidoglycan (PG) respectively, the outer membrane (OM) needs to be disrupted actively to complete lysis. A common mechanism employed by most lysis systems to overcome the OM is the use of the two-component spanins. Rz Rz1 from phage λ are the most well studied two-component spanins. Rz, the i-spanin, is an integral IM protein component with an N-terminal TMD and a highly charged C-terminal domain that extends into the periplasm. Secondary structure predictions show that the periplasmic domain of Rz has two alpha helices connected by a short linker region [52, 135]. These alpha helices have a high propensity to form coiled-coiled structures, which have been shown to be important for spanin function [52]. Rz1, the o-spanin, is an OM lipoprotein [129]. Due to an unusually high number of prolines in its periplasmic domain, Rz1 is predicted to be mostly unstructured. The two spanin subunits interact through their C-termini to form a complex that spans the entire periplasm and function to disrupt the OM [114, 115]. Previous biochemical studies have illustrated that Rz Rz1 form covalent homodimers through intermolecular disulfide bonds between cysteines at position 99, 152 in Rz and 29 in Rz1 respectively [114]. Mutational analysis substituting Cys with Ser revealed that either Cys152 in Rz or Cys29 in Rz1 is necessary and sufficient for spanin function. According to our model for Rz-Rz1 function, the spanin

complexes linking the IM and OM are assembled by the normal *sec* and *lol* pathways. In this extended form, premature function is prevented by intact PG by restricting the lateral diffusion of the complexes. Absent dissociation of the subunits, the spanin complex is confined to the local lacuna within the PG. When PG is degraded by endolysin, the spanin complex becomes active to undergo a conformational change, fusing the IM and OM. Spheroplast fusion mediated by Rz and Rz1 supports our model that spanins indeed are capable of fusion and disrupt the OM via a topological solution: fusion with the IM [117].

The unimolecular spanin gp11 of phage T1

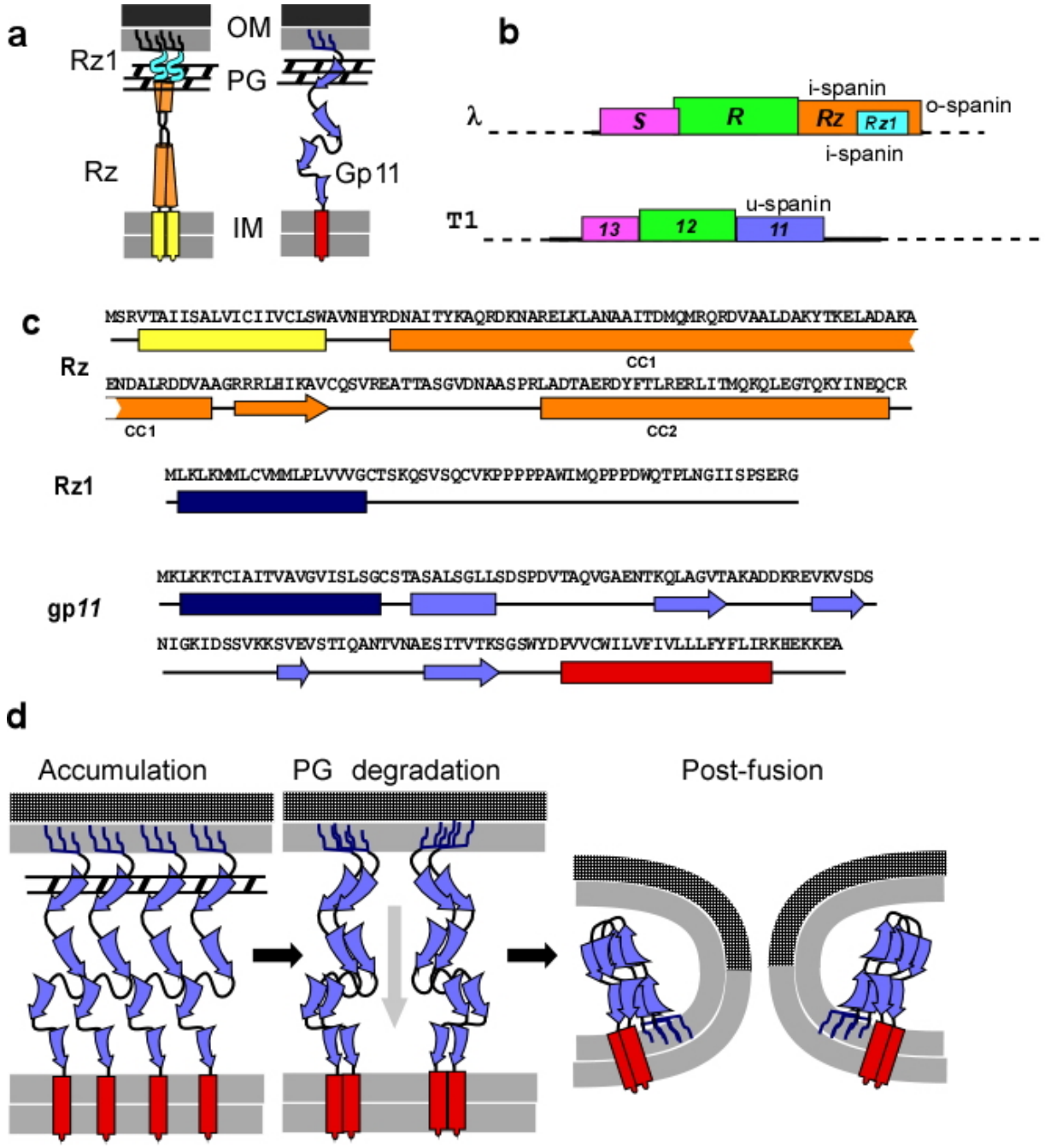
Some phages, like coliphage T1, deviate from the two-component spanins and instead encode a single component/ unimolecular spanin (u-spanin) (Fig. 4.1a) [116]. The last of the three genes in the lysis cassette of phage T1 (Fig. 4.1b), gene *11* codes for a 133 amino acid protein that has an N-terminal lipoylation signal sequence and SPaseII processing, and a C-terminal TMD (Fig. 4.1c) [116]. After post-translational processing and sorting by the Lol system, mature gp*11* is predicted to be an OM lipoprotein with a C-terminal TMD traversing the IM. This feature makes gp*11* and the other u-spanins the only class of proteins in prokaryotic biology that are physically attached to both the membranes of Gram-negative bacteria. The *λRzRz1* lysis defect can be complemented by gp*11* [116], proving that gp*11* works as a spanin to disrupt the OM.

A recent biochemical and genetic analysis gave us insights into the localization and regulation of gp*11*. It was demonstrated that both the membrane localization signals of gp*11* are necessary for function. Restricting the localization of gp*11* to the OM by deleting the C-terminal TMD (gp$11_{\Delta TMD}$) or the IM by mutating the N-terminal lipoylation Cys (gp11_{C22S}) rendered the

u-spanin non-functional. Substitution of the +2 and +3 residues (Ser23 and T24) of mature gp*11* with Asp, thus abolishing the OM sorting signal and locking the N-terminal end of gp*11* in the IM, also resulted in loss of function. All three of the above mentioned topological mutants showed a dosage-dependent, partially dominant-negative phenotype on gp*11* function. Furthermore, the C-terminal TMD of gp*11* could be entirely substituted with an artificial TMD (gp$11_{\text{art-TMD}}$) without affecting the function of gp*11*. The simplest interpretation of this result was the C-terminal TMD of gp*11* serves only to tether the gp*11* molecule to the IM and does not participate disruption of the OM.

Another important finding of this study was that the *gp11* gene has alternative start sites at position V63, V65 and active translations from either of this site resulted in the production of a short gp*11* product referred to as gp11^{a}. gp11^{a} has an "anti-spanin" effect as abolishing the start sites resulted in an earlier lysis phenotype. Also, surprisingly, inhibiting SPaseII activity using globomycin revealed that gp*11* could undergo processing by SPaseI. The product of gp*11* from SPaseI processing is thought to be the same as gene product of the gp11_{C22S} allele; localized exclusively to the IM and non-functional. In addition, manipulating the PG crosslinking using hosts auxotrophic for D-Ala showed that gp*11* caused lysis in a "holin-endolysin" independent fashion upon partial PG degradation. This also indicated that the PG meshwork acts as a negative inhibitor of gp*11* function, same as proposed for Rz-Rz1. Finally, when accumulation of gp*11* was followed over time using a C-terminal GFP fusion allele, it was observed that gp*11*-GFP accumulated in punctate foci, especially at poles and mid-cell region. Strikingly gp*11*-GFP accumulation at mid-cell showed a ring-like formation which was dependent on gp*11*-GFP being expressed in a prophage context. When gp*11*-GFP was expressed by itself, in the absence of the prophage, punctate foci phenotype was retained but the mid-cell ring like formation was not

Figure 4.1 Differences between two-component and u-spanins. a) Cartoon representation of the topology of two-component spanin from phage λ and the u-spanin from phage T1. In λ, i-spanin Rz (orange) is embedded to the IM by an N-terminal TMD (yellow rectangle) and has a periplasmic domain that constitutes two alpha helices (cylinders) connected by a linker, predicted to form coiled coils. The mature form of o-spanin Rz1 (cyan) is attached to the inner leaflet of the OM via three fatty acyl chains (dark blue lines) and has a periplasmic domain predicted to be unstructured. They interact through their C-termini to form the spanin complex, linking the IM and OM through the PG meshwork. In T1, the u-spanin gp*11* (purple) is attached to the inner leaflet of the OM, by the three fatty acyl chains (dark blue lines) at the N-terminus and to the inner membrane, through the C-terminal TMD (red rectangle). The periplasmic domain of gp*11*, predicted to be mainly extended beta sheets (arrows), connects the IM and OM through the PG meshwork. b) Lysis cassette arrangements from phages λ and T1. The boxes, color coded by pink, green, orange, cyan and purple represent the holin, endolysin, i-spanin, o-spanin and u-spanin respectively in each phage. All genes are drawn to scale and labeled with appropriate gene names. c) Predicted secondary structures of i-spanin Rz, o-spanin Rz1 from phage λ and T1 u-spanin gp*11*. The TMDs in Rz and gp*11* are shown as yellow and red rectangles respectively, while the lipoylation signal sequences in Rz1 and gp*11* are shown as dark blue rectangles. Orange rectangles and arrows indicate predicted alpha helical and beta sheet domains in Rz. Coiled-coil domains CC1 and CC2 are connected through a flexible linker region. Rz1 is predicted to be highly unstructured. Purple rectangles and arrows indicate predicted alpha helical and beta sheet domains in gp*11*. Model for gp*11* function. Gp*11* accumulates in both the inner and outer membranes of Gram-negative bacteria stuck within the PG meshwork. Once the PG is degraded by the endolysin, the gp*11* complexes can undergo higher order oligomerization and/or conformational changes along the periplasmic domain (indicated by grey arrow), fusing both the membranes together.



observed anymore. This punctate distribution was dependent on both the membrane localization signals. Both gp*11*$_{C22S}$-GFP and gp*11*$_{IRS}$-GFP exhibited a uniform distribution across the cells instead of the punctate distribution. These results indicated that gp*11* underwent oligomerization

and formed clusters within the native PG meshwork and oligomerization is dependent on gp*11* being in the extended conformation spanning the periplasm.

Model for gp11 function

Unlike the two-component spanins, gp*11* neither has predicted helical structure nor has any periplasmic cysteines for disulfide-linked oligomerization. Instead the periplasmic domain of gp*11* is predicted to be dominated by beta strands (Fig. 4.1c). Given that the u-spanin links the IM and OM, as found for the two-component spanins, the most parsimonious hypothesis is that the u-spanin also functions by IM-OM fusion (Fig. 4.1d). However, due to the differences in the secondary structures between the two-component and unimolecular spanins, it is clear that fundamental mechanistic differences exist. According to our model for gp*11* function, the gp*11* molecules accumulate in both the IM and OM, threaded through the PG lacunae, throughout the morphogenesis segment of the latent period. During this time, lateral diffusion of these molecules is restricted by the glycoside and cross-linking oligopeptide chains of the PG network. Upon destruction of the PG by the endolysin, these lacunae are broken down, setting the spanins free to diffuse laterally and oligomerize, providing free energy for conformational changes that initiate fusion of the IM and OM.

In this report, we present results of a detailed genetic analysis of the u-spanin T1gp*11*, biochemical characterization of lysis-defective gp*11* mutants and evidence that supports the model for u-spanin function, i.e. fusion of the IM and OM.

Results and discussion

Mutational analysis of gene 11

To this end, we used a medium-copy plasmid, pGp*11* reported in the previous study. This plasmid, which has gene *11* under control of the native lambda late promoter, was subjected to random mutagenesis PCR. The resultant mutant library was screened for non-functional mutants of gp*11*, as assessed by complementation of the *λRzRz1* lysis defect (Fig. 4.2a). After transformation into the *λRzRz1* lysogenic host, we tested 446 gp*11* plasmids, found 122 transformants that were defective for lysis, of which 76 could be sequenced with primers matching the original vector construction. (Table 4.1). Sequence analysis showed that 27 of these mutants contained nonsense mutations, signal sequence disruptions or frame-shifts, all of which would obviously be non-functional. Thirteen of the 41 nonsense mutants that could result from a single base change in the gp*11* reading frame were isolated, indicating that the mutational analysis was about 30% saturated. Of the rest, 21 were single missense mutants, 8 double missense and 10 triple missense mutants, and the remaining had had more than three mutations.

To further characterize the non-functional mutations identified in our screen, we focused on the single missense mutants and created them de novo in the vector pGp*11* by site directed mutagenesis. Upon testing for complementation of the *λRzRz1* lysis defect, seven of these mutations (A25E, G52R, K60E, N69K, I70F, S84L, and D106N; highlighted red in Table 4.1) turned out to be false positives. Thus, there were 14 mutants (Fig. 4.2b) in which a single amino acid change rendered gp*11* non-functional. Three of these mutations occur at the same position (D74K, D74V and D74N); thus, effectively changes in 11 of 85 positions in the periplasmic domain conferred a lysis defect. Interestingly there seems to be a requirement for a negatively

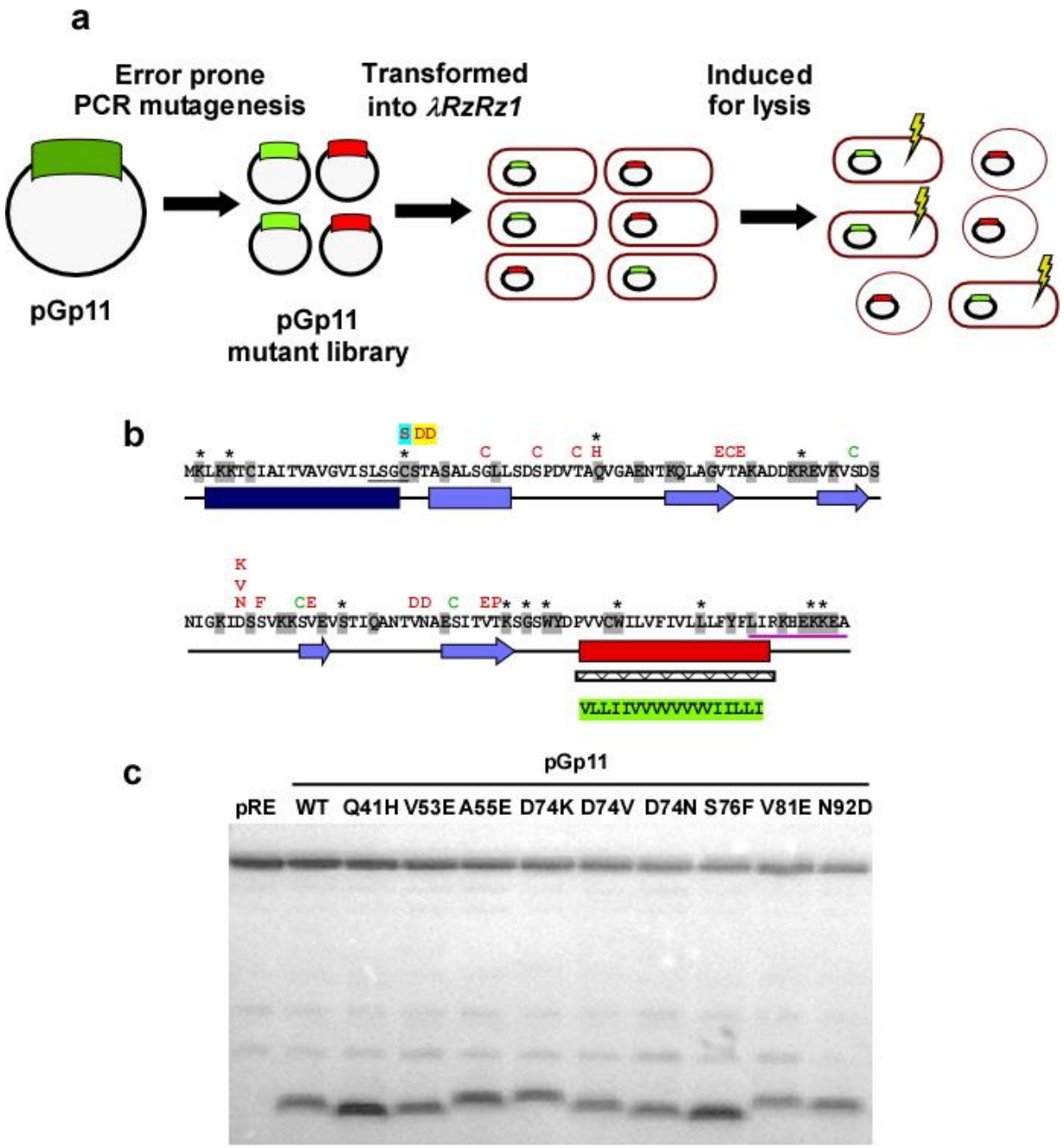
charged/basic residue at the D74 position as substitution with three different types of residues: K- positively charged, V- non-polar, and N- polar, all made gp*11* lysis-defective. Furthermore, D74V and D74N were isolated twice while the Q41H mutation was isolated three times from two mutually exclusive pools of mutant plasmids Remarkably, these positions are not clustered together or correspond to a predicted secondary structure domain but, instead, are spread throughout the periplasmic domain, suggesting the entire length of the periplasmic domain is involved in spanin function. It is possible that these positions are involved in intra or inter molecular spanin interactions that are responsible for oligomerization and/or conformational changes, which are required for spanin function.

The 14 mutants were tested for accumulation of the mutant product after induction, using immunoblotting. All the single mutants accumulated at comparable levels to gp$11_{WT,}$ but some of them showed variations in mobility on denaturing SDS gels (Fig. 4.2c). The simplest interpretation is that these mutations either change the secondary structure composition or confer conformational changes in gp*11* that could in turn affect the number of SDS-molecules bound to the protein and change the electrophoretic mobility of the protein. Three representative single missense mutants, Q41H (close to the N-terminal end), D74V (middle of the periplasmic domain) and N92D (close to the C-terminal end) were chosen for further characterization.

Visualization of gp11 mutant accumulation using gp11-GFP

We proposed that the if the mutations that made gp*11* non-functional were defective in oligomerization, we should be able to detect differences in the accumulation pattern of these mutants using our gp*11*-GFP fusion allele. To determine if the lysis-defective mutations affected

Figure 4.2 Lysis-defective mutants of gp*11*. a) Schematic showing the screening procedure for isolation of non-functional gp*11* mutants: pGp*11* was subjected to error prone mutagenesis to create a gp*11* mutant library and transformed into cells carrying spanin defective λ prophage. Individual transformants were then grown and induced for lysis. Mutations that do not affect gp*11* function (green plasmids) would be able to cause lysis. gp*11* alleles that failed to complement the *λRzRz1* lysis defect (red plasmids) were isolated and analyzed further. b) Lysis-defective mutations mapped onto gp*11* primary and predicted secondary structure: Single missense changes that render gp*11* non-functional, isolated from the mutational analysis are shown in red above the residues where the mutations occur. All the residues where a single missense mutation can result in a non-sense mutation are highlighted in grey and the nonsense mutations isolated in the screening procedure are denoted by *. Non-functional and functional cysteine changes are shown in red and green respectively. Topological mutants of gp*11* created in a previous study are also depicted: The Cys to Ser change that results in gp11_{C22S} is shown as highlighted in blue (the lipobox is underlined in black). The Asp substitutions immediately after the lipoylation cysteine that abolish the OM sorting signal in gp11_{IRS} are shown as highlighted in yellow. The hatched rectangle shows the C-terminal TMD region deleted in gp$11_{\Delta TMD}$ whereas the substituted sequence for the artificial TMD in gp$_{11art\text{-}TMD}$ is shown as highlighted in green. The pink underlined residues are the epitope site for gp*11* antibody. Expression levels of representative non-functional gp*11* alleles. Cultures of MC4100 *λRzamRz1am* carrying pGp*11* plasmids with different non-functional alleles were induced for lysis. TCA precipitates were collected from these cultures at ~40 min after induction and analyzed by western blotting. The samples in each lane are labelled at the top.



the punctate distribution, we introduced the three of the single missense changes Q41H, D74V, and N92D into the GFP fusion allele. Following the accumulation of these mutants over time, we noticed punctate distribution of GFP signal but with differences in localization of the foci. The accumulation phenotype of the Q41H allele was the closest to the functional gp*11*-GFP, showing a mid-cell ring-like accumulation (Fig 4.3 and Movie S4.1). But even with a similar localization

phenotype, these cells did not lyse. It is likely that the Q41H mutation is blocked at a later stage in the gp*11* functional pathway, i.e., probably it affects the ability of gp*11* molecules to undergo the conformational changes required to bring the IM and OM together for fusion. On the other hand, cells carrying the D74V and N92D alleles lacked the mid-cell ring-like formation. Instead, they showed accumulation in multiple foci distributed at different locations throughout the cell (Fig 4.3 and Movie S4.2, S4.3). These results suggest that the loss of function can be attributed to oligomerization defects caused by the D74V and N92D substitutions. However, the formation of GFP foci by these alleles indicates that these mutants are not completely deficient in oligomerization. It is possible that gp*11* function requires the formation of high order oligomer and these mutants are blocked at an oligomeric state lower than that required for gp*11* function.

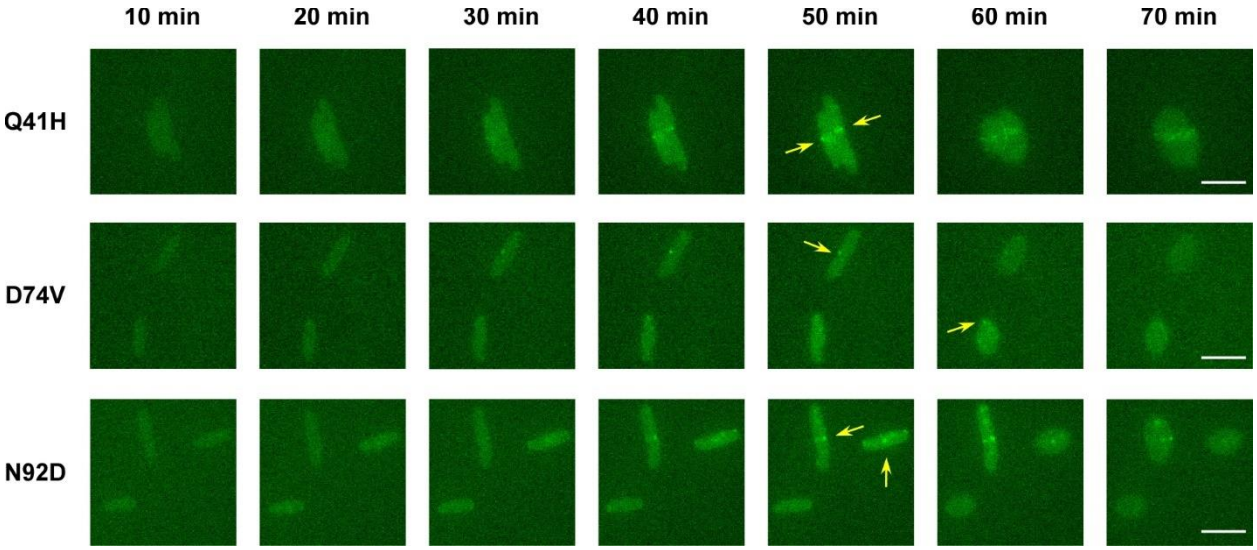


Figure 4.3 Accumulation of gp*11* lysis-defective mutants. Representative snap shots of accumulation of gp*11*-gfp non-functional mutants in thermally induced MC4100 $\lambda Rz_{am}Rz1_{am}$ cells over time. Yellow arrows indicate accumulation of gp*11*-GFP alleles as punctate foci. The scale bar (5 μM) is shown in white.

Oligomerization assay using DSP crosslinking

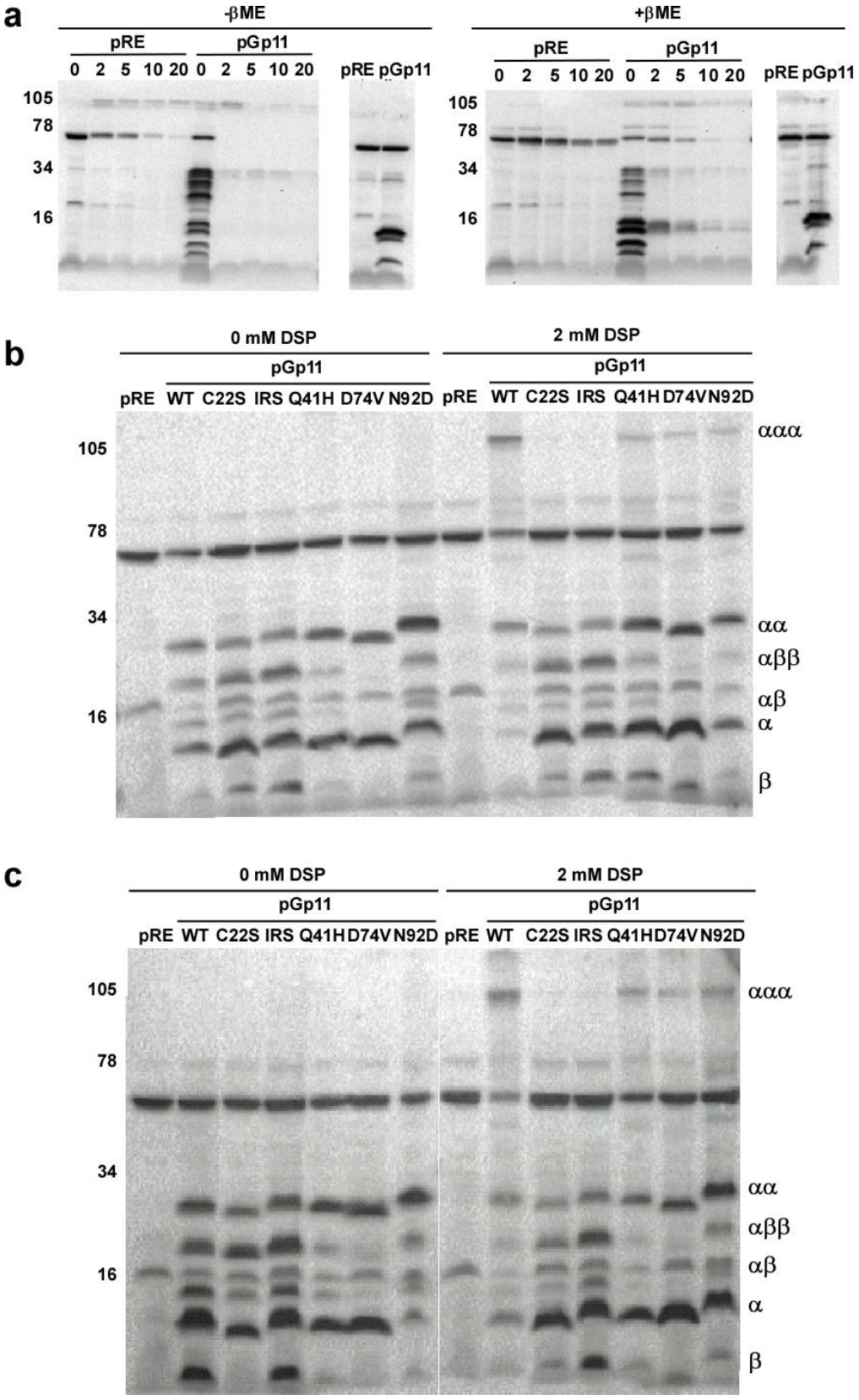
The results from the accumulation studies of non-functional gp*11*-GFP compelled us to further investigate potential defects in oligomerization in the lysis-defective mutants. To test this idea, we set up a crosslinking assay using dithio-bis-succinimidylpropionate (DSP). DSP is an amine-reactive crosslinker, having NHS-ester reactive ends and a cleavable disulfide bond in its spacer arm. It reacts with amine side chains of lysines to form disulfide linked oligomers. Cells induced for pGp*11* expression were collected and treated with different concentrations of DSP, collected by TCA precipitation, and analyzed by western blotting in either reducing or non-reducing conditions (Fig 4.4a). Surprisingly, we detected a ladder-like pattern for gp*11* even in the no crosslinker control with DMSO alone under non-reducing conditions. These results suggest that DMSO induces formation of SDS-resistant gp*11* oligomers (Fig 4.4a, left panel). Though uncommon, cases of DMSO induced protein aggregation have been reported before [222]. Nevertheless, the reduced DMSO treated samples showed a down shift of oligomer populations, indicating these oligomers were connected by a disulfide linkage. While gp*11* does not possess any periplasmic Cys, there is one Cys residue in the C-terminal TMD at position 110. Thus, the simplest explanation of this result was that the DMSO induced oligomers were linked by intermolecular disulfide linkages between the Cys*11*0 residues in neighboring TMDs. However, two lines of evidence suggested that these intermolecular disulfide bonds were probably an artefact of DMSO induced aggregation. First, it had been shown that the replacing the C-terminal TMD with an artificial TMD, which does not contain a Cys, had no effect on gp*11* function. Second, TCA precipitate of untreated gp*11* sample showed no ladder like pattern under non-reducing conditions. Thus, it is likely that the intermolecular disulfide linkages

formed by Cys110 have no functional significance. Another striking observation in the ladder like pattern is the presence of multiple species between the predicted molecular weight for a gp*11* monomer and dimer. This indicated that these species were not the result of oligomerization/aggregation of two mature gp*11* molecules. Indeed, it had been demonstrated before that expression of gp*11* results in at least two distinct translational products: mature gp*11* and gp*11*^{a}. Thus, the most plausible explanation of this observation is that these species are SDS resistant aggregates formed by different combinations of gp*11* and gp*11*a (like gp*11*:gp*11*, gp*11*:gp*11*^{a}, gp*11*^{a}:gp*11*^{a}).

Turning now to the DSP treated samples, we could not detect any bands corresponding to gp*11* in any of the DSP treated samples that were resuspended in non-reducing buffer. However, gp*11* could be detected in the reduced samples, and the amount of gp*11* detected decreased with increasing concentrations of DSP (Fig 4.4a, right panel). These results suggest that the crosslinked product consists of very high order oligomers, with a predicted molecular weight beyond the range of resolution on an SDS gel. Higher concentrations of DSP potentially create multiple disulfide linkages between gp*11* oligomers, which cannot be entirely cleaved by the reducing agent. Given the results observed with DMSO, it is possible that the oligomers formed by DSP crosslinking could also involve oligomers between gp*11* and gp*11*^{a}.

By varying the time allowed for crosslinking, we found that treatment of gp*11* with 2 mM DSP for 20 min yielded optimal results for interpretation. At these conditions, we were able to detect two significant gp*11* bands (Fig 4.4b); one corresponding to a potential gp*11* dimer (labeled with $\alpha\alpha$) and the other corresponding to a high order oligomer (labeled with $\alpha\alpha\alpha$) with a molecular weight of ~105 kDa. If it is assumed that this oligomer contains only mature gp*11*

Figure 4.4 DSP crosslinking analysis of gp*11*. a) Cells carrying the empty vector pRE or pGp*11* were induced for expression and collected after 40 min. These samples were then treated with varying concentrations of DSP in DMSO for 30 min (See materials and methods) and analyzed by western blotting under reducing and non-reducing conditions. The lanes are labeled with the concentration of DSP in each sample. TCA precipitates of untreated cells are shown on the right for each condition. b) and c) Cells carrying the empty vector pRE or different topological / non-functional pGp*11* alleles were induced for expression and collected after 40 min. These samples were then treated with 2 mM DSP in DMSO or DMSO alone for 20 min and analyzed by western blotting under non-reducing (b) and reducing conditions (c). Bands corresponding to gp11^a (β), gp*11*(α), gp*11* dimer (α), gp11^a*:* gp*11* heterodimer (αβ), gp11^a*:* gp11^a*:* gp*11* hetero-trimer (αββ) and the high order oligomer of gp*11* (ααα) are labeled.


a
-βME
+βME
pRE
pGp11
0 2 5 10 20 0 2 5 10 20
pRE pGp11
105
78
34
16
b
0 mM DSP
2 mM DSP
pGp11
pRE WT C22S IRS Q41H D74V N92D
αααα
αα
αββ
αβ
α
β
c
0 mM DSP
2 mM DSP
pGp11
pRE WT C22S IRS Q41H D74V N92D
αααα
αα
αββ
αβ
α
β

molecules, it would be predicted to be a gp*11* octamer. Nevertheless, it cannot be ruled out that this high order oligomer also contains gp11^{a} molecules.

When we subjected the topological mutants of gp*11* to the same crosslinking conditions, we noticed that the high order oligomer band was absent in both gp11_{C22S} and gp11_{IRS} (Fig 4.4b). However, oligomers that correspond to predicted dimers of gp11_{C22S} and gp11_{IRS} (labeled with $\alpha\alpha$) were still observed. This indicated that the formation of the high order oligomer was dependent on localization to both the membranes i.e. gp*11* being in the extended conformation. It also suggested that formation of the dimer form did not need the extended conformation. Furthermore, we observed bands of intermediate sizes in both the samples. According to the calculated weights, these bands were interpreted to correspond to two species: one that would be a heterodimer of gp11_{C22S}/gp11_{IRS} and gp11^{a} (labeled with $\alpha\beta$); and the other a heterotrimer consisting of one molecule of gp11_{C22S}/gp11_{IRS} and a dimer of gp11^{a} (labeled with $\alpha\beta\beta$).

Furthermore, when we subjected the three single missense mutants, Q41H, D74V, and N92D to the same crosslinking conditions, all the three mutants were able to form the high order oligomer (Fig 4.4b). However, the amount of the oligomer was significantly less compared to the gp*11* lane. There was also a notable difference in the relative population of monomers and dimers in the non-functional alleles compared to wild type gp*11*. This difference indicates that quantitative differences in oligomer formation could be the reason for loss of function in these mutants. These results are consistent with our findings with from time-lapse microscopy studies using the gp*11*-GFP allele. Interestingly, the bands corresponding to the heterodimer or heterotrimer involving gp11^{a} were present in the Q41H sample but not the D74V and N92D

samples. Since translation of gp*11*[a] starts at either V63 or V65, it is understandable that Q41H did not have an effect on the oligomer formation. These observations also further corroborate our hypothesis that D74V and N92D are defective in oligomerization.

Cysteine scanning mutagenesis

In a recent comprehensive analysis of ~600 phage genomes of Gram-negative hosts (Chapter II), we observed that nearly all two-component spanins had at least one cysteine in either the i-spanin or the o-spanin component. In contrast, none of the 58 u-spanins (Chapter II) had any periplasmic cysteines. This finding, combined with the fact that gp*11* G30C was non-functional, suggested there was an evolutionary pressure on the u-spanins to exclude periplasmic cysteines, presumably to prevent intermolecular disulfide linkages from occurring. To test this notion, we introduced cysteine mutations at six different positions across the periplasmic domain, accounting for different regions of the predicted secondary structures in the periplasmic domain (Fig. 4.2a). When we tested the ability of these Cys-substitution mutants to complement the *λRzRz1* lysis defect, we found that S35C, T39C and T54C were non-functional, whereas S66C, S80C, S95C were still able to support lysis (Fig. 4.5a). To determine whether the functionality of the mutants correlated with the presence or absence of a covalent linkage between the periplasmic cysteines, we examined samples prepared from induced cells by Western blotting (Fig. 4.5b). Surprisingly these gp*11* single cysteine mutants, irrespective of lytic function, formed intermolecular disulfide bond-linked homodimers. The efficient formation of disulfides presumably reflects the extended character of the gp*11*periplasmic domain, stretched between the two membranes such that a Cys residue in one molecule will be in an identical Z-position with the same residue of other gp*11* molecules localized in the same PG

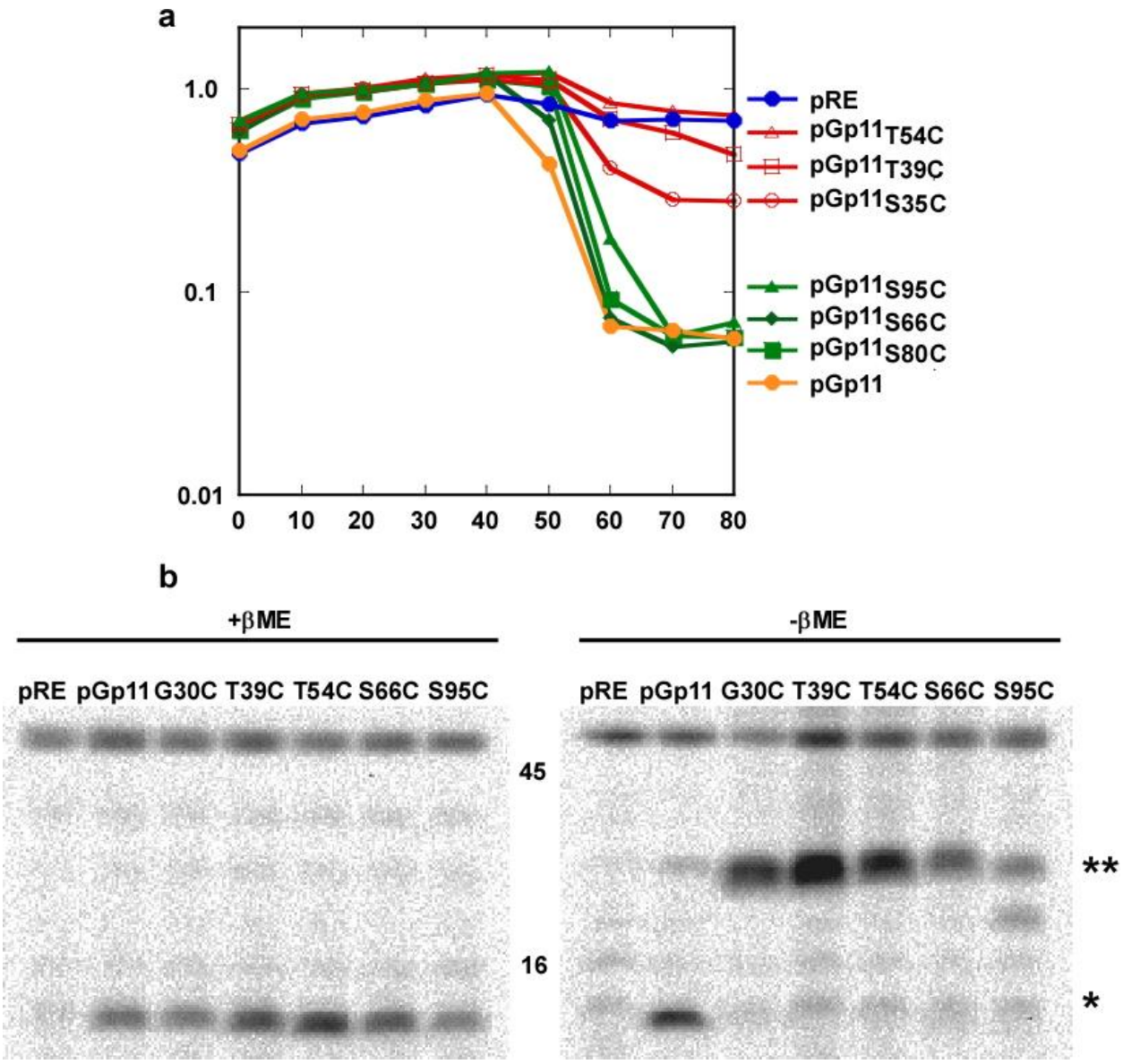


Figure 4.5 Cysteine substitution analysis of gp*11*. a) Lysis phenotype of gp*11* single cysteine mutants. Cultures of MC4100 $\lambda Rz_{am}Rz1_{am}$ carrying pRE plasmids carrying different gp*11* single cysteine mutations were grown at 30°C until A_{550} was ~0.25. The prophage and the plasmid carrying the spanin were induced by transferring to 42°C at t=0 min and moved to 37°C at t=15 min. The optical density of the culture was followed over time every 10 min. Closed blue circles = no insert (pRE vector only control), closed orange circles = gp*11*, open red circles = gp11_{S35C}, open red squares = gp11_{T39C}, open red triangles = gp11_{T54C}, closed green circles = gp11_{S66C}, closed green squares = gp11_{S80C}, closed green triangles = gp11_{S95C}. b) TCA precipitates were collected from cultures above at 50 min after induction and analyzed by western blotting. The samples in each lane are labelled at the top. The bands corresponding to the monomeric gp*11* and disulfide linked dimer are indicated by * and ** respectively.

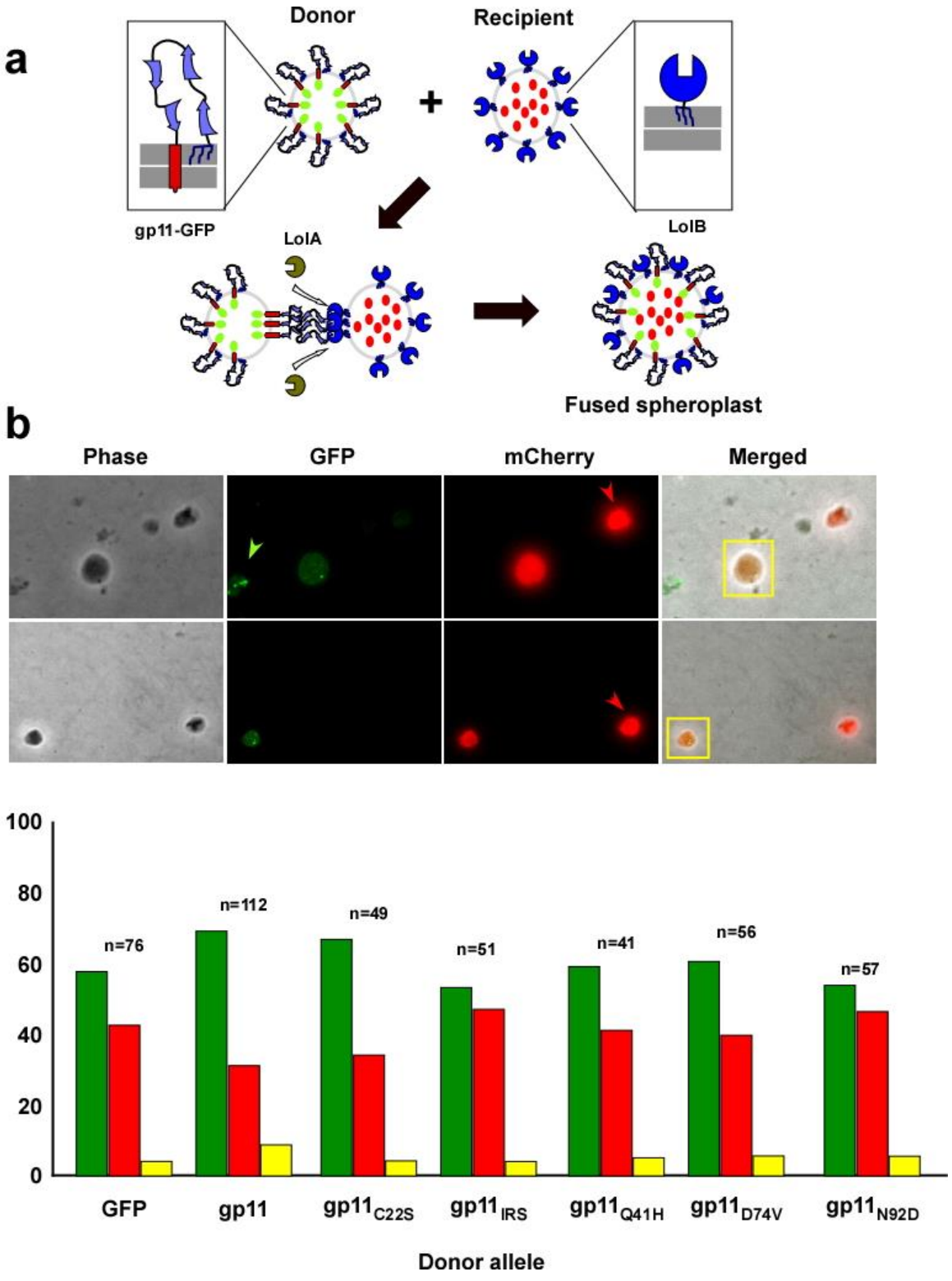
lacuna, strongly favoring Cys-Cys interactions. The fact that covalent homodimer links in the C-terminal half of the periplasmic domain are tolerated for function in the lambda system yet Cys residues are never found in nature indicates that there is a more stringent requirement in the phage contexts that feature u-spanins.

To confirm that the loss of function in the cysteine mutants was due to the intermolecular disulfide linkages and not the positions where the substitutions were introduced, we took advantage of the available *dsbA*$^-$ and *dsbC*$^-$ hosts carrying the *λRzRz1* prophage. DsbA is a thiol-disulfide oxidoreductase that catalyzes the formation of intramolecular disulfide bonds between sequential cysteine residues while DsbC is a disulfide isomerase that removes non-native disulfide bonds to favor proper folding of proteins [223]. When we tested the functionality of two non-functional (T39C and T54C) and two functional (S66C and S95C) cysteine mutants in the *dsbA*$^-$ and *dsbC*$^-$ host background to complement the λ spanin lysis defect (Table 4.2), we found that all the mutants were functional in the *dsbA*$^-$ background. On the other hand, the mutants exhibited the same phenotype in terms of functionality in the *dsbC*$^-$ background as the wild type host. These results support the notion that disulfide linkages are the reason for loss of function in the non-functional gp*11* mutants.

Gp11 supports membrane fusion

To test our model for gp11 function, we developed a spheroplast fusion assay similar to the one used for l for the λ Rz-Rz1 two-component spanin. In the fusion assay reported by Rajaure et.al [117], the spanin subunits were expressed separately in different cells, using the native allele for Rz and an Rz1 allele in which the lipoprotein was redirected to the IM by converting the two residues immediately distal to the lipoylated Cys to Asp residues.

Figure 4.6 Spheroplast fusion assay. a) Schematic representation of the spheroplast fusion assay to test the membrane fusion model for gp*11* function. The Donor and receiver spheroplasts are induced to express gp*11*-GFP and mCherry respectively. The gp*11*-GFP expressed in the donor spheroplasts accumulates in the IM, connected from both the ends. LolB in the receiver spheroplasts accumulates in the IM, making it a pseudo-OM for the Lol system. The residual diluted LolA in the spheroplast solution transfers the N-terminal lipoylated part of gp*11*-GFP to the pseudo-OM of the recipient spheroplast, which leads to the fusion of the spheroplast in the absence of the regulating PG lacunae. b) Phase, GFP only, mCherry only and merged fluorescence images of the mixture of donor and receiver spheroplasts. Green and red arrows indicate donor and recipient spheroplasts while the fused spheroplasts are highlighted in yellow boxes. c) Histogram showing the percentage of fused and non-fused spheroplasts with respect to the gp*11* allele used in the donor spheroplasts. Green, red and yellow bars indicate the percentage of GFP labeled, mCherry labeled and fused spheroplasts respectively. The gp*11* allele expressed in the donor spheroplast is indicated below. The bars corresponding to the GFP sample indicate the control with no spanin. All the experiments had recipient spheroplasts expressing mCherry. The number of total spheroplasts counted for each sample is mentioned above each set of bars.


a
Donor
Recipient
gp11-GFP
LolA
LolB
Fused spheroplast
b
Phase
GFP
mCherry
Merged
c
% of spheroplasts
100
80
60
40
20
0
n=76
n=112
n=49
n=51
n=41
n=56
n=57
GFP
gp11
gp11C22S
gp11IRS
gp11Q41H
gp11D74V
gp11N92D
Donor allele

Spheroplasts were then prepared from these cells, exposing the periplasmic domains. Spheroplast formation from mixtures of these cells resulted efficient formation of fused cells, in a manner dependent on the lytic functionality of the periplasmic domains of each component. This indicated that Rz-Rz1 were capable of fusing two lipid bilayers.

To modify this assay for gp*11*, we used two different fluorescently labelled spheroplasts, designated as donor and recipient spheroplasts. The donor and recipient spheroplasts are prepared from cells carrying the pGp*11*-GFP and pmCherry plasmid (mCherry under the control of the pR' promoter) respectively Spheroplasts are prepared by treating cells with EDTA and lysozyme, which destabilize the OM and degrade the PG respectively. Since gp*11* is connected to both the membranes and causes lysis in the absence of PG, it is not possible to prepare spheroplasts after accumulation of gp*11*. Upon induction of pR' expression in the donor spheroplasts, gp*11* is expected to accumulate in the IM, attached by both the N- and C-terminal ends. The outer membrane receiver for lipoprotein delivery LolB itself is an OM lipoprotein. Thus, any newly synthesized LolB in the recipient spheroplasts is also expected to accumulate in the IM after spheroplasting. As a result, a recipient spheroplast carrying LolB molecules can act as "pseudo OM" for lipoprotein delivery. Upon mixing the donor and recipient spheroplasts, we hypothesize that the residual LolA in the spheroplast solution should transport the lipoylated N-terminus of gp*11* from the donor to the recipient spheroplasts. This spanin connected to two IMs essentially imitates the spanin linking the IM and OM in the host cell. According to our hypothesis, since there is no PG to inhibit gp*11* function, gp*11* molecules should fuse the donor and recipient spheroplasts.

Accordingly, when the donor and recipient spheroplasts were mixed at a 1:1 ratio and induced for expression of gp*11* along with the fluorescent markers, we found ~ 9% of the mixed

spheroplasts were labeled with both GFP and mCherry signals, compared to ~4% for controls lacking gp*11*. This efficiency is comparable to the 10% fusion efficiency reported for Rz-Rz1. The doubly labeled spheroplasts were larger in size compared to the spheroplasts with only GFP or mCherry signal. When we expressed the topological mutants gp11_{C22S} and gp11_{IRS} in the donor spheroplasts, fusion efficiency was reduced to background. This indicated that the fusion event is dependent on gp*11* being connected to both the membranes, supporting our previous results that both the membrane localization signals of gp*11* are necessary for lytic function. Furthermore, we tested the fusion efficiency of the three non-functional alleles of gp*11*, Q41H, D74V and N92D. All these non-functional alleles showed a fusion efficiency significantly lower than that of the functional gp*11*, but still higher than background. These results suggest that the lysis-defective mutants are impaired for the fusion activity. However, spheroplast fusion appears to be less demanding than the fusion of IM and OM, as these mutants are totally non-functional in the whole cell context.

Conclusion and next steps

Here we present results of the first detailed mutational analysis done on the u-spanin T1gp*11*. We were able to isolate 14 mutations at 11 positions spread throughout the periplasmic domain of gp*11* that make gp*11* non-functional. That indicates that only 12% of the 85 residues in the periplasmic domain are non-permissive to changes. Given that these mutations do not cluster together, it seems that the entire length of periplasmic domain is a determining factor for gp11 function. This interpretation would mean that we should be able to isolate more mutations along the length of the periplasmic domain that would also confer a lysis defect. Given that only 30% saturation is achieved, may be parameters used for the error prone PCR mutagenesis in

creating the gp*11* mutant library can be modified to increase the saturation of the genetic analysis. Furthermore, the mutant libraries can be screened for conditional mutants such as temperature sensitive mutants that could give us more insight into gp*11* function. Also, a suppressor hunt can be implemented using the currently available non-functional mutants which could give us clues about potential intra and intermolecular interactions.

We hypothesize that the lysis-defective mutants are potentially blocked at different stages of the gp*11* function pathway. Biochemical evidence indicated that some of these mutants exhibit oligomerization defects and are potentially blocked from attaining the required concentration of a high order oligomer that is required for gp*11* function. Despite the insights provided by the current crosslinking assay reported here, the results must be interpreted with caution. Firstly, the no-crosslinker control, DMSO, showed formation of SDS resistant oligomers. These oligomers seem to be covalently linked through intermolecular disulfide bonds through the Cys110 residue that is buried in the IM, which has been shown to be irrelevant for gp*11* function. Second, the presence of multiple translational products of the *gp11* transcript complicates the analysis of the crosslinked species. It is possible that repeating the crosslinking experiments with a construct of gp*11* that lacks both the alternative start sites as well as Cys110 will give a better insight into the oligomerization levels of gp*11*. Given the anti-spanin effect of gp*11*^{a}, it cannot be ruled out that the interactions revealed here are functionally relevant. Nevertheless, the results of the crosslinking assay raise a few questions such as a) What is the composition of the high order oligomer that is required for gp*11* function? b) Which interactions are affected by the mutations conferring the lysis defect; the ones between homo-oligomers of gp*11* or hetero-oligomers involving gp*11* and gp*11*^{a}? Oligomerization defects in the lysis-defective mutations were also seen in the subcellular localization studies using gp*11*-GFP.

Cysteine substitution analysis gave us contrasting results for the mutations introducing a Cys residue in the N-terminal half compared to the C-terminal half. Interestingly, all these single cysteine mutants formed homodimers linked through intermolecular disulfide bonds. It was intriguing to us that all of the gp*11* population existed in the dimer form as detected under non-reducing conditions. For the effective formation of disulfide linkages at such level, it is likely that the Cys residues in neighboring gp*11* molecules must be within a close distance of each other. Extending this interpretation, it can be thought that residues at multiple positions along the Z-axis of extended gp*11* molecules localized in the same PG lacuna are involved in "inter-spanin" interactions. These interactions potentially form oligomers of low order, probably dimers or trimers, within a single lacuna. The gp*11* molecules in these low order oligomers would thus be primed for higher order oligomerization once the PG meshwork is degraded by the endolysin. Further experiments need to be done to test whether the covalent linkages restrict further oligomerization by constraining the freedom of orientation of gp*11* molecules in the dimer.

The results from the spheroplast fusion assay indicate gp*11* can mediate membrane fusion in spheroplasts. These observations support our idea that gp*11* achieves OM disruption, the last step in lysis of Gram-negative hosts, by fusing the OM with the IM. Based on the similarities with class II viral fusion proteins in terms of secondary structure, it can be hypothesized that membrane fusion proceeds through a condensation event of the gp*11* periplasmic domain. The extended conformation of gp*11* molecule in the presence of intact PG represents the prefusion structure. The degradation of PG allows for a transition from intermolecular to intramolecular interactions between the beta-sheets, thus condensing the periplasmic domain and bringing the IM and OM together. In the post-fusion conformation, it would be expected that the N-terminal lipoylated end and the C-terminal TMD end up together. The calculated efficiency of our

spheroplast fusion assay is likely an underestimate of the ability of gp*11* to mediate membrane fusion. For example, a "selfing" event in which two donor spheroplasts fuse together would not result in doubly labeled spheroplasts. Also, we would not be able to detect a fusion event between a donor spheroplast and an already fused, doubly labeled fused spheroplast. Nevertheless, the reduction in fusion efficiency for the lysis-defective alleles of gp*11* validates the effectiveness of the spheroplast fusion assay to assess gp*11* function. Future work should focus on refining the conditions for the fusion assay to increase the efficiency for functional spanins as well as decreasing the background signal. Addition of purified LolA instead of relying on the diluted residual LolA in the spheroplast solution could lead to a more efficient transfer of the N-terminal end of gp*11* from the donor to recipient spheroplasts. In addition, overexpression of a missense allele of LolB, where the OM sorting signal is abolished, to increase the concentration of LolB in the recipient spheroplasts can also be implemented.

Materials and methods

Bacterial strains, phages, plasmids, and growth conditions

The bacterial strains, bacteriophages, and plasmids used in this study are described in Table 4.3. Bacterial cultures were grown in standard LB medium, supplemented with antibiotics; ampicillin (Amp, 100 μg/ml) and kanamycin (Kan, 40 100 μg/ml) as appropriate. Expression of plasmids carrying genes under the lambda late promoter pR' was induced by providing the antiterminator Q in trans from either the induced prophage or the p*Q* plasmid. For experiments to test complementation of the *λRzRz1* lysis defect, MC4100 (*λ*900$Rz_{am}Rz1_{am}$) cells were transformed with the pGp*11* plasmid carrying the concerned allele and transformants selected on Amp-Kan plates at 30°C were used to inoculate overnight cultures. The overnight culture was diluted 1:250 into 25 ml fresh LB media and grown to an OD of ~0.25 at 30°C. Lysis was thermally induced by a shift from 30°C to 42°C for 15 min, followed by continued growth at 37°C. $MgCl_2$ was added to a final concentration of 10 mM at the time of induction to stabilize the OM. Growth and lysis of cultures over time was monitored by A_{550} using the Gilford Stasar III spectrophotometer. For expression of pGp*11* alleles in a non-lysogenic context, MG1655 *ΔtonA* $lacI^{q1}$ p*Q* cells were transformed with the pGp*11*-plasmid carrying the concerned allele and transformants selected on Amp-Kan plates at 37°C were used to inoculate overnight cultures. The overnight culture was diluted 1:250 into 25 ml fresh LB media and grown to an OD of ~0.2 before inducing by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to the final concentration of 1 mM for plasmid induction. Lysis curves were plotted using Kaleidagraph.

Isolation and conformation of lysis-defective mutations in gp11

Three different gp*11* mutant libraries were created by subjecting the pGp*11* plasmid to error prone PCR mutagenesis using the GeneMorph II Random Mutagenesis kit. PCR reaction conditions were manipulated such that mutation frequency varied between 0-9 mutations/kb. Mutagenized pGp*11* plasmids were restriction digested using BamHI^{HF} and HindIII^{HF} and the inserts carrying the gp*11* mutant alleles were purified using the QIAquick gel extraction kit. The inserts were ligated back into a clean pRE vector backbone digested with the same restriction enzymes. The vector backbone was dephosphorylated using the Antarctic phosphatase from NEB to prevent the backbone from self-ligating. The entire overnight ligation product was transformed into XL1Blue competent cells. Ten random colonies were chosen to determine the efficiency of mutagenesis. Once the optimal mutagenesis efficiency was confirmed, a pool of mutant plasmids was prepared from a slurry of colonies from the transformation plate. 2-5 µl of the mutant library plasmid sample was transformed into MC4100 (λ900$Rz_{am}Rz1_{am}$) cells and isolated single colonies were picked to inoculate two sets of 5 ml cultures supplemented with 10mM $MgCl_2$, grown at 30º Celsius. One set of the cultures was thermally induced for lysis and samples showing turbidity after an hour were declared to be lysis-defective. Plasmids were extracted from corresponding overnight cultures from the second set of cultures and sent for sequencing. The sequencing results were analyzed through the Sequencher program. Once the mutations were identified, primers were designed to introduce them into pGp*11* using site-directed mutagenesis and confirmed for the lysis defect.

SDS-PAGE and western blotting

SDS-PAGE and western blotting were performed as described previously [105]. TCA pellets of whole cell samples from cultures expressing the different pGp*11* alleles were acetone washed, resuspended in 1X SDS-PAGE buffer with β-mercaptoethanol (100mM BME). Protein samples were loaded onto 10% resolving Tris-tricine polyacrylamide gels after loading volumes were normalized according to A_{550} at the time of collection of TCA precipitates. Proteins were transferred to PVDF membrane (Pall Life Sciences) using a Hoefer TE unit at 0.5 mA for 2 hours. Antibodies (Genscript) were generated in rabbits against a synthetic peptide CIRKHEKKEA, corresponding to the C-terminal end of gp*11*. The primary antibodies were used at a dilution of 1:1000 while the secondary antibody, goat-anti-rabbit-HRP (Thermo Scientific), was used at a dilution of 1:5000. Chemiluminescence was detected using a Bio-Rad XR Gel Doc system. SeeBlue Plus2 (Invitrogen) pre-stained standard served as a molecular mass standard.

Fluorescence microscopy

MC4100 (*λ*900$Rz_{am}Rz1_{am}$) cells carrying the different lysis-defective alleles of pGp*11-gfp* were grown to an OD of ~0.25 at 30°C. Lysis was thermally induced by a shift to 42°C for 15 min, and then 1 μL of sample was placed onto a coverslip (24 × 50 mm). After gently covering the sample with an agarose pad (1% wt/vol in LB+ 10 mM $MgCl_2$) of ~5 × 5 mm dimensions, the cells were continued to grow at 37°C while imaging was performed simultaneously on an inverted microscope (Ti-E, Nikon, Tokyo, Japan) with a cage incubator (InVivo Scientific, St. Louis, MO) set at 37°C. Images were taken using 100× objective (Plan Fluo, NA 1.40, oil immersion) with standard filter sets and a cooled EMCCD camera (iXon 3 897, Andor, Belfast,

United Kingdom). Images were taken every 1 min for the specified times through the phase-contrast and GFP channels (100 ms exposure) over 8-10 different coordinates. Image analysis and processing was done using NIS-Elements imaging software.

DSP Crosslinking assay

MG1655 *ΔtonA lacI*^{q1} p*Q* cells expressing the different allelic versions of pGp*11* were induced for expression by the addition of IPTG. 2ml of each culture was harvested at A_{550} of ~ 0.7-0.8 and washed with 1ml PBS buffer (pH7.2) twice. Cells were resuspended in 640µl PBS. 20µl of 5mM DSP in DMSO was added to the cells and allowed to incubate for 5 minutes at RT. 16µl of 1M Tris buffer (pH7.5) was then added to the reaction and allowed to incubate 15min at RT. The reaction was then pelleted by centrifugation at 14k rpm for 10min, and the supernatant was discarded. The protein pellet was then washed twice with 1ml of ice cold acetone and air dried overnight.

Spheroplast fusion assay and fluorescence microscopy

MG1655 *ΔtonA lacI*^{q1} cells carrying the p*Q* and pGp*11-gfp* plasmids were used to prepare the donor spheroplasts while MG1655 *ΔtonA lacI*^{q1} cells carrying the p*Q* and pmCherry plasmids were used to prepare the recipient spheroplasts. For control experiments, cells carrying pGFP, instead of pGp*11-gfp* were used as donors. Similarly, pGp*11-gfp* plasmids carrying the lysis-defective mutations were used in donor cells to test their fusion efficiency. 10 mL of separately grown stationary phase donor and recipient strain cultures were mixed and harvested by centrifugation at 5000 x g for 10 min at 4°C using a Thermo Scientific F15S-8x50cy rotor to prepare spheroplasts as described before [136]. 1 ml of the spheroplasts were mixed with 500 µl

of 4X LB, 200 μl of 20% glucose, 300 μl of 45% sucrose and induced with 1 mM IPTG for 2-3 hours at 37°C to express gp*11*-GFP and mCherry. 1μL of sample was placed onto a 24x50mm coverslip and gently covered with agarose pad (0.6% wt/vol in 0.25M sucrose) and immediately imaged as described above for the fluorescence spectroscopy. For each experiment, snapshots of at least 10 different frames were captured. Image analysis was done using the NIS-Elements imaging software on each frame individually and number of spheroplasts emitting red, green, or both signals were manually scored. The percentage of fused spheroplast was calculated using a basic formula [% = (number of fused cells/ total)*100; total = red + green + both].

Tables

Table 4.1 List of lysis-defective mutants of gp*11*

S.No	Mutations	Remarks
1	A25E	single
2	G30C	single
3	Q41H (isolated 2 times)	single
4	Q41H (K100)	single
5	G52R	single
6	V53E	single
7	A55E	single
8	K60E	single
9	N69K	single
10	I70F	single
11	D74N (isolated twice)	single
12	D74V (isolated twice)	single
13	D74K	single
14	S76F (L118)	single
15	V81E	single
16	S84L	single
17	V91D	single
18	N92D (isolated twice)	single
19	V98E	single
20	T99P	single
21	D106N	single
22	L3R, D74V	double
23	G15V, D74G (P107)	double
24	A25P, A55E	double
25	D37A, A93D	double
26	A40E, T54A	double
27	V53E, K72R	double
28	T54I, V81E	double
29	D74V, V81E	double
30	V16M, G30D, G71A	triple
31	T24S, R61Q, K127N	triple
32	A25V, Q49R, L120M	triple
33	S26F, S95G, F115I (K100)	triple
34	L32F, Q87P, N92D	triple
35	V42D, L50V, S66R	triple

Table 4.1 (continued)

S.No	Mutations	Remarks
36	A55G, K100I, G102R	triple
37	N69K, D106N, I116M (I125)	triple
38	G71S, S103R, L118P	triple
39	V91D, P107H, I115M	triple
40	I10V, D67G, I86N, P107S	4 mutations
41	V12D, I86S, A93D, Y105N (I17)	4 mutations
42	A25E, V53E, E94G, V98L (I112, K127)	4 mutations
43	L32P, D58Y, V81G, S83L (L19, G71, D74)	4 mutations
44	K60E, G71S, S103R, L118P	4 mutations
45	P36L, Q41H, I96I, S104N, L119F (A40)	5 mutations
46	C7G, L28R, Q41H, E82D, V109G, C110F (L118, K127)	6 mutations
47	Q41* (G15)	nonsense mutant
48	S84*	nonsense mutant
49	G102* (isolated twice)	nonsense mutant
50	W111* (isolated twice)	nonsense mutant
51	L119* (S33)	nonsense mutant
52	K130*	nonsense mutant
53	K2*, G30D	nonsense mutant
54	S18F, G102* (G71)	nonsense mutant
55	C22*, S26F	nonsense mutant
56	W104*, F115Y	nonsense mutant
57	W111*, I125V	nonsense mutant
58	K5*, I70T, V98L	nonsense mutant
59	S20T, S29N, R61*	nonsense mutant
60	D74Y, K100*, P106T, Y120#	nonsense mutant
61	T6A, D74V, T99P, F121Y, K130*	nonsense mutant
62	K60E, D67G, T85S, N88S, N91Y, I116M, K131* (L3, A27, S33, V77)	nonsense mutant
63	L31# (H128, K131)	frameshift
64	V114# (S33)	frameshift
65	V38#, R61Q (V42)	frameshift
66	G21R, A56#	frameshift (lipobox disrupted)

Table 4.1 (continued)

S.No	Mutations	Remarks
67	S68#, F123L (D74, T85)	frameshift
68	C22R, S26F	lipoylation Cys mutated
69	C22F, D37G, S76F (I8, A57)	lipoylation Cys mutated
70	L19P, A55G	lipobox disrupted
71	G21R, T54A (G102)	lipobox disrupted
72	G21D, Q41H, V81A (V12)	lipobox disrupted

(X) – silent mutations ; * - nonsense mutation ; # - frame shift due to insertion/deletion. The single missense mutants that were isolated in the screen that turned out to be false positives are highlighted in red.

Table 4.2 Lysis phenotype of gp11 cysteine mutants in *dsbA*$^-$ and *dsbC*$^-$ hosts ()

	Host background		
Mutant	Wild type	*dsbA*$^-$	*dsbC*$^-$
T39C	-	+	-
T54C	-	+	-
S66C	+	+	+
S95C	+	+	+

+ means functional and – means lysis-defective

Table 4.3 Phages, bacterial strains and plasmids

Name	Description	Reference
Phages		
λ900	*λΔ(stf tfa)::cat cI857 bor::kan*; carries Cam^R and Kan^R	Laboratory stock
λ900Rz_{am} $Rz1_{am}$	λ900 carrying the Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles of the spanin genes	[115]
***E. coli* strains**		
MC4100 *tonA::Tn10*	Escherichia coli K-12 *F araD139 Δ(argF-lac)U169 rpsL15 relA1 flbB3501 deo pstF25 rbsR tonA*	Laboratory stock
MC4100 (*λ*900$Rz_{am}Rz1_{am}$)	MC4100 *tonA::Tn10* carrying λ900 prophage with Rz_{Q100am} $Rz1_{W38am}$ nonsense alleles	[115]
MG1655 *ΔtonA* $lacI^{q1}$ p*Q*	MG1655 *ΔtonA* $lacI^{q1}$ carrying the p*Q* plasmid	[221]
Plasmids		
pRE	Medium copy vector with the λ late promoter pR' that is transcriptionally activated by λQ; Amp^R	[221]
pGp*11*	pRE plasmid carrying the *T1gp11* gene	[116]
pGp11_{C22S}	pRE plasmid carrying the *T1gp11* allele in which the lipoylation Cys22 residue is substituted with Ser	Laboratory stock
pGp11_{IRS}	pRE plasmid carrying the *T1gp11* allele in which the Ser23 and Thr24 residues are substituted with Asp	Laboratory stock
pGp11_X	pGp*11* plasmid carrying the lysis-defective alleles of gp*11* or the cysteine substitutions	This study
pGp*11-gfp*	pRE plasmid carrying the *T1gp11* allele with a translational GFP fusion at the C-terminus	Laboratory stock
pGp$11\text{-}gfp_X$	pGp11-gfp plasmid carrying the lysis-defective alleles of gp*11* where X is Q41H, D74V, N92D	This study
p*Q*	λ *Q* gene cloned under $P_{lac/ara-1}$ promoter in a low copy number plasmid pZS-24; Kan^R	[106]
pmCherry	mCherry cloned downstream of the pR' promoter	[117]

CHAPTER V

DISCUSSION CONCLUSION, AND NEXT STEPS

The work presented in this dissertation provides significant insights into the last step of Gram-negative host lysis by phages, disruption of the OM. To restate the general premise, phages employ either two-component spanins or unimolecular spanins to disrupt the OM barrier [116]. Two-component spanins like Rz-Rz1 from phage λ consist of an integral inner membrane protein: i-spanin, and an outer membrane lipoprotein: o-spanin, that form a complex spanning the periplasm. Two-component spanins exist in three different genetic architectures; embedded, overlapped and separated. In contrast, the unimolecular spanins, like gp*11* from phage T1, have an N-terminal lipoylation signal sequence and a C-terminal transmembrane domain to account for the topology requirements. Our proposed model for spanin function, for both spanin types, follows a common theme: the spanins cause fusion of the OM with the IM, resulting in direct exposure of the cytoplasm to the environment and thus an osmotic catastrophe.

Identification, classification, and characterization of spanins

In chapter II, we present a SpaninDB (Table S2.1) consisting of 528 two-component spanins and 58 unimolecular spanins identified in genomes of dsDNA phages infecting Gram-negative hosts. While comprehensive bioinformatic analyses have been done on phage holins [224] and endolysins [225] before, this is the first study on spanins at such a large scale. The inefficiency of automated annotation attempts and the subsequent success using a manual search protocol in creating the SpaninDB highlighted the difficulty associated with annotation of spanins. This could be mainly attributed to the inability of current gene-calling programs to

recognize the uncommon architectures found in the two-component spanin systems. This strongly suggests the need for gene-calling programs that consider uncommon genetic architectures for future phage annotation pipelines. This study also lays the groundwork for future bioinformatic investigations on lysis proteins. Using a threshold of 40% sequence identity over 40% sequence length, we were able to group the spanins into 143 i-spanin, 125 o-spanin and 13 u-spanin families (Table 2.2, 2.3, 2.4). Interestingly, more than 40% of these families from each type were singletons, emphasizing the extreme diversity in this class of lysis proteins. The difference in composition of i- and o-spanin families sheds new light on their independent evolutionary paths. The most striking aspect of this observation is that the two-component spanins of the embedded and overlapped architecture can evolve independently even while sharing the same DNA sequence. Additionally, pair-wise sequence alignments of two-component spanins that belonged to the same family, but different genetic architecture, allowed us to interpret the evolutionary dynamics of spanin gene arrangement (Fig. 2.6). The simplest interpretation is that the two-component spanins transitioned from the separated architecture to the overlapped or embedded architecture through a series of duplication, modification and deletion events.

The bioinformatics survey also served to expand our knowledge about the u-spanins. One of the most significant findings from this study was that u-spanins were found to be associated with both “canonical holin-cytoplasmic endolysin” as well as “pinholin-SAR endolysin” systems. Each of the three steps in lysis of Gram-negative hosts by phages has the choice of two fundamentally different options: hole formation in the IM by the holin or the pinholin, PG degradation by the cytoplasmic endolysin or SAR endolysin, and then OM disruption by the two-component spanin complex or the u-spanin. Only six of the eight possible combinations would

be functionally compatible since the pinholins cannot accomplish the release of cytoplasmic endolysins. The seven u-spanins reported by Summer et.al, [116] were associated with only with "pinholin-SAR endolysin systems", giving rise to a notion that u-spanins could be found only in these systems. Not only do our findings dismiss this idea but also complete the identification of all these six possible combinations. Multiple sequence alignments of the gp*11* u-spanin family revealed that all family members followed a similar secondary structure organization even with different periplasmic domain lengths (Fig. 2.9). This observation that u-spanins maintain a conserved beta-sheet domain organization is in accordance with our proposed model for u-spanin function involving condensation of beta-sheet segments.

Primary structure analysis revealed significant differences in the secondary structure compositions for the periplasmic domains of the two-component and unimolecular spanin types (Table S2.2). Variations were observed even within the three different genetic architectures of the two-component spanins. Furthermore, multiple sequence alignments of the periplasmic domains of two-component spanins provided two major insights. First, alignments of sequences belonging to the same family demonstrated conserved secondary structure patterns and domain organization within family members (Fig. 2.3, 2.4, 2.5). Second, these predicted structural features were overwhelmingly distinct for different families and architectures. Taken together, these findings suggest different spanins utilize different mechanisms to achieve the same result, i.e. fusion of IM and OM. Especially, the range of alpha-helical and beta-sheet secondary structure combinations found in spanins has not been seen before in fusion proteins. Biochemical and structural studies on spanins with these different compositions, in terms of prefusion and postfusion conformations, would greatly add to the existing diversity in membrane fusion mechanisms.

Along with the identification of new spanins in more than 580 genomes, the bioinformatic search also failed to identify spanins in ~80 phages with our criteria. While it is possible these phages might not face as robust an OM barrier in their natural environment compared to the lab conditions, other modes of OM disruption cannot be ruled out. There has been some progress through research in our lab to elucidate such alternate mechanisms of OM disruption. We have been able to identify at least one protein, gp28 from phage KT, that has shown OM disruption activity. gp28 is a highly basic 56 aa long protein with no signal sequence or TMD and very few identifiable homologs. Designated as a "disruptin", it is not clear yet how gp28 recognizes and overcomes the OM barrier. Nevertheless, this preliminary result hints at other possible routes of countering the OM employed by phages.

Translation of gp11 produces multiple products with different localization

Chapter III presents a detailed investigation into the localization and regulation of the u-spanin gp*11*. Using topological mutants localizing to only one of the membranes, we demonstrated that both the predicted membrane localization signals of gp*11* were true and necessary for function (Fig. 3.2). Furthermore, we showed that substituting the entire C-terminal TMD of gp*11* with an artificial TMD (gp$11_{\text{art-TMD}}$) did not affect the functionality of the u-spanin (Fig. 3.2). This result is in accordance with the findings of Berry et.al, where replacing the TMD of Rz with the TMD of FtsI did not affect Rz-Rz1 function. Together, these results indicate that the role of TMDs in both the two-component and u-spanin systems is only to physically anchor the spanins to the IM. This observation that the sequence specificity of the TMD anchor being irrelevant for function has been seen in other fusion systems as well. Experiments involving

membrane separation are in progress to establish gp*11* as the only protein to be experimentally shown to accumulate in both the membranes of Gram-negative bacteria as a single polypeptide molecule. Nevertheless, it is safe to infer this conclusion from the results of the topological mutants.

A surprising revelation from this study was that gp*11* could be processed by both SPaseI and II, as suggested by two different lines of evidence. First, the *gp11*$_{C22S}$ allele produced a lower molecular weight product instead of the expected unprocessed product (Fig. 3.2d). Second, inhibiting SPaseII processing by treatment with globomycin also resulted in the accumulation of a lower molecular weight product instead of the unprocessed product (Fig. 3.3b). The most plausible explanation for these results is that the lower molecular weight product in both the cases is essentially the same and is a result of SPaseI processing. Due to the similarities in signal sequences specific for SPaseI and SPaseII, most lipoproteins are predicted to also have a SPaseI processing site [206]. The presence of a conserved lipobox motif at the end of the signal sequence is the differentiating feature between the two signal sequences. However, , this is the first time a lipoprotein has been shown to actively processed by both the signal peptidases. Nonetheless, the resultant product from SPaseI processing would be non-functional as it would not be tethered to the OM. In addition, this study also showed that active translation from downstream start sites results in the production of a shorter gp*11* product, designated as gp*11*^{a} (Fig 3.2d). In comparison to mature gp*11*, gp*11*^{a} contains only the C-terminal half and is tethered to the IM through the C-terminal TMD. In summary, translation of the *gp11* transcript results in three different products. Only the mature gp*11* processed by SpaseII is connected to both the membranes and is functional while the other two accumulate solely in the IM and are non-functional.

Gp11 is regulated at multiple levels

The work presented in chapter III also gave us insights into the regulation of gp*11* function. Abolishing the downstream start site from which gp*11*^{a} is translated resulted in an earlier lysis phenotype (Fig. 3.3c). This indicated that gp*11*^{a} exerted a negative dominant effect on gp*11* function, and serves as an intrinsic "anti-spanin", establishing the first level of regulation. We also demonstrated that all the topological mutants exhibited a dosage-dependent partial negative dominant effect on gp*11* activity (Fig 3.5a). Extrapolating this result, it can be thought the gp*11* product from SPaseI processing, which we predict to be the same as the gp*11*$_{C22S}$ product, would have a similar inhibitory effect. Thus, the competitive processing of gp*11* between the SPases acts as the second level of regulation of u-spanin function. Furthermore, investigating the effect of partial PG degradation on gp*11* function using host racemase mutants revealed that PG acts as a negative inhibitor of gp*11* function These results are consistent with the previously reported observations and interpretations for Rz-Rz1 [117]. After post-translational processing and sorting, it is expected that mature gp*11* molecules would be trapped in the PG meshwork. The PG crosslinks would not allow gp*11* molecules confined to a lacuna to freely diffuse and/or interact with gp*11* molecules from other lacunae, thus establishing a third level of regulation. Currently it is unclear how large of a gap in PG crosslinks is sufficient for spanins to cause membrane fusion. Also, the relevance of gp*11*^{a} and SPaseI processing needs to be tested in the context of a T1 phage infection. Nevertheless, at least in the context of working in trans to spanin-less λ prophage, gp*11* is regulated at three different levels as stated above. Thus, apart from the holin regulation by antiholin [125], and the regulation of SAR

endolysins by disulfide caging [226] and/or membrane tethering [227], this presents an additional and novel instance of lysis protein regulation.

Time-lapse fluorescence microscopy indicates clustered assembly of gp11

The first clues for gp*11* oligomerization came from our studies using a functional C-terminal GFP fusion allele of gp*11* presented in chapter III. Following the accumulation of gp*11* expressed in trans to a spanin-less prophage over time, we observed gp*11* accumulated in punctate spots (Fig. 3.6b, Mov. S3.1). Interestingly, the majority of gp*11*-GFP was clustered at the poles and in a ring-like formation at the mid-cell region. Interestingly, the mid-cell accumulation was dependent on being expressed in the lambda prophage context, suggesting it is likely related to the expression of the λ *kil* gene. Inhibition of FtsZ activity by Kil could lead to local PG remodeling, leading to the formation of larger lacunae at mid-cell region. The increased diameter would allow for more spanin molecules to localize within a lacuna at the mid-cell, which could cause the ring-like formation. It is also possible that inhibition of FtsZ forces the cell to redirect its secretion machinery to mid-cell to account for the changed dynamics of PG turnover. The localized concentration of the transportation machinery could increase the number of gp*11* molecules being threaded through the lacunae at mid-cell region and cause the ring-like accumulation. The functional relevance of this mid-cell accumulation of gp*11* needs to be tested in the context of a T1 phage infection. Nevertheless, the punctate distribution is indicative of gp*11* clusters accumulating within the PG lacunae. The cluster formation requires the extended conformation of gp*11* as explained by the lack of punctate foci in the topological versions of the gp*11*-GFP allele that localized exclusively to the IM (Fig. 3.6c,d, Mov. S3.2, 3.3). It is likely that periplasmic interactions between the beta sheet structures of adjacent gp*11* molecules within the

same mesh-hole of the PG network could lead to gp*11* oligomers and thus the punctate GFP foci. It can be hypothesized that local oligomeric concentrations are needed to set up or prime the u-spanins for rapid OM disruption once the PG is degraded.

Lysis-defective mutants of gp11 are blocked at different stages of gp11 function

The genetic analysis of gp*11* using random PCR mutagenesis presented in chapter IV yielded 14 single missense mutations at 11 positions that caused a lysis defect. All these mutations are localized to the periplasmic domain, thus should not have any effect on gp*11* maturation. This would mean that these mutations are blocked in the gp*11* functional pathway at a stage downstream of the maturation step. Based on our model for gp*11* function, we propose these mutants are defective in oligomerization and/or undergoing local/global conformational rearrangements that are required for spanin function. Differences in oligomerization patterns in the DSP crosslinking assay and subcellular localization studies indicate there are at least two different levels at which function is obstructed in these mutants. With a saturation of only 30% reached, future work should focus on isolating and characterizing more lysis-defective mutations from the random mutant library. Another approach to increase the size of the non-functional mutant pool would be to implement site directed mutagenesis, taking into consideration the observations from the bioinformatic analysis. Multiple sequence alignment of the T1gp*11* family revealed that multiple positions across the periplasmic domain were highly conserved. Indeed, two of the three positions chosen for further analysis in chapter IV: Q41 and D74 are conserved in all the homologs. Together, these observations emphasize the importance of these residues at

those positions for spanin function. Thus, using a site directed mutagenesis approach targeting other highly conserved positions could shed more light on the functionally important domains.

Apart from the traditional mutant hunt, isolation of conditional mutants could also provide deeper insights into spanin function. that could give us more insight into gp*11* function. The availability of host racemase mutants provides us with a platform to manipulate the level of PG crosslinking through the supply of the PG precursor D-Ala. It would be interesting to test if we can isolate conditional mutants that can effect "holin-endolysin" independent lysis only below a certain threshold of PG crosslinking. Also, a suppressor hunt can be implemented using the currently available non-functional mutants which could give us clues about the molecular basis for the lysis defect caused by the original mutation. Coincidentally, of the only few available such studies done on membrane fusion proteins, one focused on the alphavirus E1 protein, a class II viral membrane fusion protein [228]. Their results indicate second site suppressors promote the conformational change steps required for function, either by better domain-domain packing and/or by reducing the steric hindrances. Thus, similar studies on the u-spanin could help us identify and understand the role of intramolecular and intermolecular domain-domain interactions.

There also exists a need for a better way to detect oligomerization defects in the lysis-defective mutants. A yeast two-hybrid or bacterial two-hybrid system involving systematic domain deletions from the N-terminal and C-terminal end of the periplasmic domain could be go ahead step in this direction. This would help us detect not only the local interactions between adjacent domains but also global interactions between far away domains within the same molecule. This could also give us clues about the prefusion and postfusion structures of the u-spanin and the necessary conformational changes required for the transition between the two

stages. Lysis-defective mutations can then be introduced into these input alleles to classify mutants based on what interactions they are deficient in. Eventually, efforts need to be made in the direction of working out the solution and/or crystal structure of gp*11* , in both the prefusion and postfusion modes, to get a better idea of the effects of these mutations at the molecular level.

Periplasmic cysteines have opposing effect on two-component and unimolecular spanins

It had already been demonstrated that all the three cysteines in the λ Rz-Rz1 spanin complex, Cys99 and 152 in Rz and Cys29 in Rz1, formed intermolecular disulfide bonds [129]. At least one of these covalent linkages at either Cys152 in Rz or Cys29 in Rz1 was required for Rz-Rz1 function. To further corroborate the role of intermolecular disulfide bonds in two-component spanin function, the bioinformatic analysis in chapter II revealed that 96% of the two-component spanins identified in our study possessed at least one periplasmic cysteine in either of the two subunits (Fig. 2.7a, Table S2.5a). Moreover, a positional analysis indicated that spanin complexes tend to maintain the periplasmic cysteines close to the heterotypic interaction site of the two subunits (Fig. 2.7d, Table S2.5e). Hence, it could be conceivably hypothesized that the covalent linkages stabilize the heterotetrameric spanin complex. Multiple sequence alignments of spanin families also indicated that periplasmic cysteines were highly conserved across members (Fig. 2.3, 2.4, 2.5). In addition, the necessity of periplasmic cysteines for spanin function was substantiated by our genetic and biochemical analysis of the T4 two-component spanins PseT.3 and PseT.2. Our work illustrated that both the cysteines in PseT.2, Cys87 and Cys98, were involved in intermolecular disulfide bond formation (Fig. 2.8). However, in contrast to the findings reported with Rz-Rz1, both the cysteines in PseT.2 were required for lysis. Taking the data from western blots of the PseT.2 cysteine mutants into consideration, a possible

explanation for the requirement of both the cysteines is that the maturation pathway for the T4 spanin complex is different from that of the λ spanins. It can be proposed that maturation proceeds through a potential isomerization step of intramolecular disulfide bonds to intermolecular linkages. This idea can be further extended to all the two-component spanins given the considerable variation in the number of periplasmic cysteines for different spanins. It is possible that, in addition to their proposed role in stabilization of spanin complexes, periplasmic cysteines play a role in proper maturation of the spanin complex. It is imaginable that spanins with multiple periplasmic cysteines could even potentially regulate function by manipulating the populations linked through intramolecular and intermolecular disulfide bonds.

On the other hand, all the 58 u-spanins identified in chapter II lacked any periplasmic cysteines. Apart from the predicted secondary structure differences, this could be considered as the second major feature that differentiated the periplasmic domains of u-spanins from that of two-component spanins. This observation not only hinted at a possible evolutionary pressure on the u-spanins to avoid periplasmic cysteines, but also once more emphasized the mechanistic differences in mode of action between two-component and u-spanin types. Furthermore, a lysis-defective mutation introducing a single cysteine into the periplasmic domain of gp*11*, G30C, was isolated in our non-functional mutant search reported in chapter IV (Fig. 4.2). This observation was consistent with our interpretation from the bioinformatic analysis. In addition, a mutational analysis introducing cysteines into 6 different positions across the periplasmic domain of gp*11*, served as test to our hypothesis about periplasmic cysteines in u-spanins. Interestingly, three of these mutants, S35C, T39C, and T54C turned out to be non-functional, while the others, S66C, S80C, and S95C still retained functionality (Fig. 4.5a). Irrespective of the effect on gp*11* function, all the cysteine mutants formed homodimers linked by intermolecular disulfide bonds

(Fig. 4.5b). The formation of these disulfide bonds was dependent on DsbA and function was restored in the non-functional cysteine mutants in a *dsbA*$^-$ background (Table 4.2). These results indicate that covalent linkages in the N-terminal half of gp*11* confer a lysis defect but are permissive in the C-terminal half. However, these results need to be interpreted with caution as the bioinformatics survey indicates a more stringent constraint on periplasmic cysteines. The functionality of the cysteine substitutions should be tested in the context of a T1 phage infection to approve our interpretation about periplasmic cysteines and the intermolecular linkages imposed by them.

The role of the host *dsb* system in the formation of intramolecular disulfide bonds has been extensively studied [73, 223]. However, apart from our knowledge about the role of DsbA and DsbC in homodimerization of Rz and Rz1 [129], not much is known about the mechanism of intermolecular disulfide bond formation in the periplasm. Very few instances of intermolecular disulfide bonds in prokaryotes have been reported so far [229-232]. Of these, only the VirB7 and VirB9 proteins of the *A. tumefaciens* T-complex transport apparatus form intermolecular linkages that are functionally significant. VirB7 forms disulfide-linked homodimers and heterodimers with VirB9, both of which are necessary for the assembly of the transport complex. Given the functional relevance of such linkages in the two-component spanins, they could serve as strong subject to study the intermolecular disulfide bond formation processes as well as disulfide bond isomerization in Gram-negatives.

Gp11 fuses membranes to achieve OM disruption

The results from the spheroplast fusion assay presented in chapter IV indicate that gp*11* is capable of membrane fusion. These results strongly support the notion that gp*11*, just like Rz-

Rz1, makes use of a topological solution to overcome the OM barrier, fusing it with the IM. Despite the promising results, the high background signal from the current design of the spheroplast fusion assay calls for a better means of assaying the fusogenic activity of gp*11*. The existing scheme of the fusion assay depends on expression of gp*11* from spheroplasts, action of the diluted residual LolA in the spheroplast solution and accumulation of LolB in the recipient spheroplasts for fusion. Further, it also depends on the detectable levels of fluorescent markers in the spheroplasts to score successful fusion events. Suggestions were made in chapter IV to increase the efficiency of the spheroplast fusion through addition of purified LolA to the spheroplast solution and redirecting LolB to the IM in recipient spheroplasts. Changes can also be implemented in the method of detecting fusion. For example, a fluorescence resonance energy transfer (FRET) based approach can be used to study the mixing of spheroplasts in the presence of gp*11*. The use of single molecule FRET to observe real-time fusion reactions, especially the ones mediated by the SNARE proteins is well documented [233]. Thus, instead of relying on expression of fluorescent markers in spheroplasts, this method would rely on differential lipid labeling of the IMs of the donor and recipient spheroplasts. An other option could be the use of the recently developed technique of single particle burst analysis spectroscopy (BAS) [234]. BAS has been shown to be capable of sensitive quantitative membrane dynamics measurements and detect changes in sizes and concentrations of liposomes. One of the major problems that we face in the current scheme of detection of fusion the spheroplast fusion assay is the inherent fragile nature of the spheroplasts. To hold the spheroplasts in place for fluorescence microscopy, a thin layer of agarose gel is applied on the top. How much ever thin, the agarose cover applies a great amount of physical stress on the spheroplasts and causes them to burst, thus effectively being unfavorable to measurements. Use of BAS would allow us to bypass the need for

fluorescence markers and microscopy entirely as the measurements would be based on the sizes of the spheroplasts in the solution. Indeed, it has been observed in the spheroplast fusion assay, although not quantified, that the doubly labeled fused spheroplasts are larger in size compared to spheroplasts showing a single fluorescence signal. This would also abolish any diminishing effects of external stress or forces on the calculation of the fusion efficiency. Furthermore, an *in vitro* assay using optical tweezers can be designed to study the secondary structure dynamics in the periplasmic domain of gp*11* [235, 236]. The ability to identify and characterize multiple intermediate conformations using the optical tweezers could be of great advantage in deciphering the various stages of gp*11* function pathway.

REFERENCES

1. Hendrix RW. Bacteriophages: Evolution of the Majority. Theoretical Population Biology 2002, 61(4):471-480.

2. Suttle CA. Marine viruses - major players in the global ecosystem. Nature Reviews Microbiology 2007, 5:801.

3. Kimura M, Jia Z-J, Nakayama N, Asakawa S. Ecology of viruses in soils: Past, present and future perspectives. Soil Science and Plant Nutrition 2008, 54(1):1-32.

4. Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. Research in Microbiology 2008, 159(5):349-357.

5. Sime-Ngando T. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. Frontiers in Microbiology 2014, 5:355.

6. Scanlan PD. Bacteria-Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. Trends in Microbiology, 25(8):614-623.

7. Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D *et al*. Raw Sewage Harbors Diverse Viral Populations. mBio 2011, 2(5).

8. Prestel E, Salamitou S, DuBow MS. An examination of the bacteriophages and bacteria of the Namib desert. The Journal of Microbiology 2008, 46(4):364.

9. Suttle CA. Viruses in the sea. Nature 2005, 437:356.

10. Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F. Phage Community Dynamics in Hot Springs. Applied and Environmental Microbiology 2004, 70(3):1633-1640.

11. Filippova SN, Surgucheva NA, Sorokin VV, Akimov VN, Karnysheva EA, Brushkov AV, Andersen D, Gal'chenko VF. Bacteriophages in Arctic and Antarctic low-temperature systems. Microbiology 2016, 85(3):359-366.

12. Koskella B, Brockhurst Michael A. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiology Reviews 2014, 38(5):916-931.

13. Middelboe M, Hagström A, Blackburn N, Sinn B, Fischer U, Borch NH, Pinhassi J, Simu K, Lorenz MG. Effects of Bacteriophages on the Population Dynamics ofFour Strains of Pelagic Marine Bacteria. Microbial Ecology 2001, 42(3):395-406.

14. Clokie MRJ, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage 2011, 1(1):31-45.

15. Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, Weinbauer M. Major viral impact on the functioning of benthic deep-sea ecosystems. Nature 2008, 454:1084.

16. Weinbauer Markus G, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? Environmental Microbiology 2003, 6(1):1-11.

17. Smith DL, James CE, Sergeant MJ, Yaxian Y, Saunders JR, McCarthy AJ, Allison HE. Short-Tailed Stx Phages Exploit the Conserved YaeT Protein To Disseminate Shiga Toxin Genes among Enterobacteria. Journal of Bacteriology 2007, 189(20):7223-7233.

18. Abedon ST, LeJeune JT. Why Bacteriophage Encode Exotoxins and other Virulence Factors. Evolutionary Bioinformatics Online 2005, 1:97-110.

19. Hatfull GF, Hendrix RW. Bacteriophages and their Genomes. Current opinion in virology 2011, 1(4):298-303.

20. Belcaid M, Bergeron A, Poisson G. Mosaic Graphs and Comparative Genomics in Phage Communities. Journal of Computational Biology 2010, 17(9):1315-1326.

21. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR *et al*. Origins of Highly Mosaic Mycobacteriophage Genomes. Cell 2003, 113(2):171-182.

22. Hankin M. L'action bactéricide des eaux de la Jumna et du Gange sur le vibrion du choléra. Ann Inst Pasteur (Paris) 1896, 10:511–523.

23. Twort FW. An investigation on the nature of ultra-microscopic viruses. The Lancet 1915, 186(4814):1241-1243.

24. D'Herelle F. Sur un microbe invisible antagonistic des bacilles dysenterique. Compt Rend Acad Sci Paris 1917, 165:373-375.

25. Duckworth DH. "Who discovered bacteriophage?". Bacteriological Reviews 1976, 40(4):793-802.

26. Sulakvelidze A, Alavidze Z, Morris JG. Bacteriophage Therapy. Antimicrobial Agents and Chemotherapy 2001, 45(3):649-659.

27. Schlesinger M. Zur frage der chemischen zusammensetzung des bakteriophagen. Biochem Z 1934, 273:306-311.

28. Luria SE, Anderson TF. The Identification and Characterization of Bacteriophages with the Electron Microscope. Proceedings of the National Academy of Sciences of the United States of America 1942, 28(4):127-130.121.

29. Fischer EP. Max Delbrück. Genetics 2007, 177(2):673-676.

30. Susman M. The Cold Spring Harbor Phage Course (1945-1970): A 50th Anniversary Remembrance. Genetics 1995, 139(3):1101-1106.

31. Watanabe R, Matsumoto T, Sano G, Ishii Y, Tateda K, Sumiyama Y, Uchiyama J, Sakurai S, Matsuzaki S, Imai S *et al*. Efficacy of Bacteriophage Therapy against Gut-Derived Sepsis Caused by Pseudomonas aeruginosa in Mice. Antimicrobial Agents and Chemotherapy 2007, 51(2):446-452.

32. Pouillot F, Chomton M, Blois H, Courroux C, Noelig J, Bidet P, Bingen E, Bonacorsi S. Efficacy of Bacteriophage Therapy in Experimental Sepsis and Meningitis Caused by a Clone O25b:H4-ST131 Escherichia coli Strain Producing CTX-M-15. Antimicrobial Agents and Chemotherapy 2012, 56(7):3568-3575.

33. Jun JW, Shin TH, Kim JH, Shin SP, Han JE, Heo GJ, De Zoysa M, Shin GW, Chai JY, Park SC. Bacteriophage Therapy of a Vibrio parahaemolyticus Infection Caused by a Multiple-Antibiotic–Resistant O3:K6 Pandemic Clinical Strain. The Journal of Infectious Diseases 2014, 210(1):72-78.

34. Wills QF, Kerrigan C, Soothill JS. Experimental Bacteriophage Protection against Staphylococcus aureus Abscesses in a Rabbit Model. Antimicrobial Agents and Chemotherapy 2005, 49(3):1220-1221.

35. LaVergne S, Hamilton T, Biswas B, Kumaraswamy M, Schooley RT, Wooten D. Phage Therapy for a Multidrug-Resistant Acinetobacter baumannii Craniectomy Site Infection. Open Forum Infectious Diseases 2018, 5(4):ofy064.

36. King AMQ AM, Carstens EB, Lefkowitz EJ, editors. . Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier/Academic Press; London, United Kingdom 2011.

37. Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D. Hyperexpansion of RNA Bacteriophage Diversity. PLoS Biology 2016, 14(3):e1002409.

38. Maniloff J, Ackermann HW. Taxonomy of bacterial viruses: establishment of tailed virus genera and the other Caudovirales. Archives of Virology 1998, 143(10):2051-2063.

39. Veesler D, Cambillau C. A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries. Microbiology and Molecular Biology Reviews 2011, 75(3):423-433.

40. Blakely G. A Genetic Switch, Third Edition, Phage Lambda Revisited. M. Ptashne, vol. 84; 2004.

41. Little JW. Lysogeny, Prophage Induction, and Lysogenic Conversion. In: *Phages.* American Society of Microbiology; 2005.

42. Bertani G. Lysogeny. In: *Advances in Virus Research.* Edited by Smith KM, Lauffer MA, vol. 5: Academic Press; 1958: 151-193.

43. Nanda AM, Thormann K, Frunzke J. Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. Journal of Bacteriology 2015, 197(3):410-419.

44. Silhavy TJ, Kahne D, Walker S. The Bacterial Cell Envelope. Cold Spring Harbor Perspectives in Biology 2010, 2(5):a000414.

45. Gram C. Uber die isolirte Farbung der Schizomyceten in Schnitt- and Trockenpraparaten. . Fortschritte der Medizin 1884, 2:185-189.

46. Ingram LO. Changes in lipid composition of Escherichia coli resulting from growth with organic solvents and with food additives. Applied and Environmental Microbiology 1977, 33(5):1233-1236.

47. Oursel D, Loutelier-Bourhis C, Orange N, Chevalier S, Norris V, Lange Catherine M. Lipid composition of membranes of Escherichia coli by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. Rapid Communications in Mass Spectrometry 2007, 21(11):1721-1728.

48. Papanastasiou M, Orfanoudaki G, Koukaki M, Kountourakis N, Sardis MF, Aivaliotis M, Karamanou S, Economou A. The Escherichia coli Peripheral Inner Membrane Proteome. Molecular & Cellular Proteomics : MCP 2013, 12(3):599-610.

49. Matias VRF, Al-Amoudi A, Dubochet J, Beveridge TJ. Cryo-Transmission Electron Microscopy of Frozen-Hydrated Sections of Escherichia coli and Pseudomonas aeruginosa. Journal of Bacteriology 2003, 185(20):6112-6118.

50. Seltmann G, Holst O. Periplasmic Space and Rigid Layer. In: *The Bacterial Cell Wall.* Edited by Seltmann G, Holst O. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002: 103-132.

51. Miller SI, Salama NR. The gram-negative bacterial periplasm: Size matters. PLOS Biology 2018, 16(1):e2004935.

52. Berry J, Savva C, Holzenburg A, Young R. The lambda spanin components Rz and Rz1 undergo tertiary and quaternary rearrangements upon complex formation. Protein Science : A Publication of the Protein Society 2010, 19(10):1967-1977.

53. Cayley DS, Guttman HJ, Record MT. Biophysical characterization of changes in amounts and activity of Escherichia coli cell and compartment water and turgor pressure in response to osmotic stress. Biophysical Journal 2000, 78(4):1748-1764.

54. Höltje J-V. Growth of the Stress-Bearing and Shape-Maintaining Murein Sacculus of Escherichia coli. Microbiology and Molecular Biology Reviews 1998, 62(1):181-203.

55. Harz H, Burgdorf K, Höltje J-V. Isolation and separation of the glycan strands from murein of Escherichia coli by reversed-phase high-performance liquid chromatography. Analytical Biochemistry 1990, 190(1):120-128.

56. Vollmer W, Blanot D, De Pedro Miguel A. Peptidoglycan structure and architecture. FEMS Microbiology Reviews 2008, 32(2):149-167.

57. Heppel LA. Selective Release of Enzymes from Bacteria. Science 1967, 156(3781):1451.

58. Kennedy EP, Rumley MK. Osmotic regulation of biosynthesis of membrane-derived oligosaccharides in Escherichia coli. Journal of Bacteriology 1988, 170(6):2457-2461.

59. Kennedy EP. Osmotic regulation and the biosynthesis of membrane-derived oligosaccharides in Escherichia coli. Proceedings of the National Academy of Sciences 1982, 79(4):1092.

60. Bayer ME. Ultrastructure and organization of the bacterial envelope. Annals of the New York Academy of Sciences 1974, 235(1):6-28.

61. Leive L. THE BARRIER FUNCTION OF THE GRAM-NEGATIVE ENVELOPE. Annals of the New York Academy of Sciences 1974, 235(1):109-129.

62. Schindler M, Osborn MJ. Interaction of divalent cations and polymyxin B with lipopolysaccharide. Biochemistry 1979, 18(20):4425-4430.

63. Braun V, Wolff H. The Murein-Lipoprotein Linkage in the Cell Wall of Escherichia coli. European Journal of Biochemistry 1970, 14(2):387-391.

64. Nikaido H. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiology and Molecular Biology Reviews 2003, 67(4):593-656.

65. Park JT. Cytoplasmic membrane. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Edited by Neidhardt FC. Washington, D.C.: American Society for Microbiology; 1996: pp. 58-87.

66. Nikaido H. Outer membrane. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Edited by Neidhardt FC. Washington, D.C.: American Society for Microbiology; 1996: pp. 29-47.

67. Driessen AJM, Nouwen N. Protein Translocation Across the Bacterial Cytoplasmic Membrane. Annual Review of Biochemistry 2008, 77(1):643-667.

68. Berg Bvd, Clemons Jr WM, Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA. X-ray structure of a protein-conducting channel. Nature 2003, 427:36.

69. Rapoport TA. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature 2007, 450:663.

70. Zimmer J, Nam Y, Rapoport TA. Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 2008, 455:936.

71. Randall LH, S. SecB, one small chaperone in the complex milieu of the cell. CMLS, Cell Mol Life Sci 2002, 59(1617).

72. Paetzel M, Karla A, Strynadka NCJ, Dalbey RE. Signal Peptidases. Chemical Reviews 2002, 102(12):4549-4580.

73. Inaba K. Disulfide Bond Formation System in Escherichia coli. The Journal of Biochemistry 2009, 146(5):591-597.

74. Patel R, Smith SM, Robinson C. Protein transport by the bacterial Tat pathway. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 2014, 1843(8):1620-1628.

75. Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. Biochimica et Biophysica Acta (BBA) - Biomembranes 2008, 1778(9):1735-1756.

76. Tokuda H, Matsuyama S-i. Sorting of lipoproteins to the outer membrane in E. coli. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 2004, 1693(1):5-13.

77. Okuda S, Tokuda H. Lipoprotein Sorting in Bacteria. Annual Review of Microbiology 2011, 65(1):239-259.

78. Narita S, Tanaka K, Matsuyama S, Tokuda H. Disruption of lolCDE, encoding an ATP-binding cassette transporter, is lethal for Escherichia coli and prevents release of lipoproteins from the inner membrane. J Bacteriol 2002, 184(5):1417-1422.

79. Tajima T, Yokota N, Matsuyama S-i, Tokuda H. Genetic analyses of the in vivo function of LolA, a periplasmic chaperone involved in the outer membrane localization of Escherichia coli lipoproteins. FEBS Letters 1998, 439(1–2):51-54.

80. Tanaka K, Matsuyama SI, Tokuda H. Deletion of lolB, encoding an outer membrane lipoprotein, is lethal for Escherichia coli and causes accumulation of lipoprotein localization intermediates in the periplasm. J Bacteriol 2001, 183(22):6538-6542.

81. Tsukahara J, Narita S-i, Tokuda H. Real time analysis of lipoprotein transfer from LolA to LolB by means of surface plasmon resonance. FEBS Letters 2009, 583(18):2987-2990.

82. Hara T, Matsuyama S-i, Tokuda H. Mechanism Underlying the Inner Membrane Retention of Escherichia coli Lipoproteins Caused by Lol Avoidance Signals. Journal of Biological Chemistry 2003, 278(41):40408-40414.

83. Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 2014, 4:e28281.

84. Storms ZJ, Sauvageau D. Modeling tailed bacteriophage adsorption: Insight into mechanisms. Virology 2015, 485:355-362.

85. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiology Letters 2016, 363(4):fnw002.

86. Ackermann HW. Bacteriophage observations and evolution. Research in Microbiology 2003, 154(4):245-251.

87. David IF, Max G. Lytic Mode of Lambda Development. Cold Spring Harbor Monograph Archive; Volume 13 (1983): Lambda II 1983.

88. Echols H. Chapter 13 Regulation of Lytic Development. Cold Spring Harbor Monograph Archive; Volume 02 (1971): The Bacteriophage Lambda 1971.

89. Thomas R. Regulation of Gene Expression in Bacteriophage Lambda. In: *Current Topics in Microbiology and Immunology / Ergebnisse der Mikrobiologie und Immunitätsforschung: 1971// 1971; Berlin, Heidelberg*: Springer Berlin Heidelberg; 1971: 13-42.

90. Wang I-N. Lysis Timing and Bacteriophage Fitness. Genetics 2006, 172(1):17-26.

91. Chang CY, Nam K, Young R. S gene expression and the timing of lysis by bacteriophage lambda. Journal of Bacteriology 1995, 177(11):3283-3294.

92. Wang I-N, Smith DL, Young R. Holins: The Protein Clocks of Bacteriophage Infections. Annual Review of Microbiology 2000, 54(1):799-825.

93. Abedon ST. Lysis of lysis-inhibited bacteriophage T4-infected cells. Journal of Bacteriology 1992, 174(24):8073-8080.

94. Paddison P, Abedon ST, Dressman HK, Gailbreath K, Tracy J, Mosser E, Neitzel J, Guttman B, Kutter E. The roles of the bacteriophage T4 r genes in lysis inhibition and fine-structure genetics: a new perspective. Genetics 1998, 148(4):1539-1550.

95. Moussa SH, Kuznetsov V, Tran TAT, Sacchettini JC, Young R. Protein determinants of phage T4 lysis inhibition. Protein Science : A Publication of the Protein Society 2012, 21(4):571-582.

96. Wagner Patrick L, Livny J, Neely Melody N, Acheson David WK, Friedman David I, Waldor Matthew K. Bacteriophage control of Shiga toxin 1 production and release by Escherichia coli. Molecular Microbiology 2002, 44(4):957-970.

97. Krüger A, Lucchesi PMA. Shiga toxins and stx phages: highly diverse entities. Microbiology 2015, 161(3):451-462.

98. Balcazar JL. Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. PLoS Pathogens 2014, 10(7):e1004219.

99. Brown-Jaque M, Calero-Cáceres W, Muniesa M. Transfer of antibiotic-resistance genes via phage-related mobile elements. Plasmid 2015, 79:1-7.

100. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology 2015, 23:171-178.

101. Young R. Bacteriophage lysis: mechanism and regulation. Microbiological Reviews 1992, 56(3):430-481.

102. Young R. Phage lysis: three steps, three choices, one outcome. J Microbiol 2014, 52(3):243-258.

103. Wang I-N, Deaton J, Young R. Sizing the Holin Lesion with an Endolysin-β-Galactosidase Fusion. Journal of Bacteriology 2003, 185(3):779-787.

104. White R, Chiba S, Pang T, Dewey JS, Savva CG, Holzenburg A, Pogliano K, Young R. Holin triggering in real time. Proceedings of the National Academy of Sciences 2011, 108(2):798-803.

105. Gründling A, Bläsi U, Young R. Genetic and Biochemical Analysis of Dimer and Oligomer Interactions of the λ S Holin. Journal of Bacteriology 2000, 182(21):6082-6090.

106. Gründling A, Manson MD, Young R. Holins kill without warning. Proceedings of the National Academy of Sciences of the United States of America 2001, 98(16):9348-9352.

107. Savva CG, Dewey JS, Moussa SH, To KH, Holzenburg A, Young R. Stable micron-scale holes are a general feature of canonical holins. Molecular microbiology 2014, 91(1):57-65.

108. Dewey JS, Savva CG, White RL, Vitha S, Holzenburg A, Young R. Micron-scale holes terminate the phage infection cycle. Proceedings of the National Academy of Sciences 2010, 107(5):2219-2223.

109. Xu M, Struck DK, Deaton J, Wang IN, Young R. The signal arrest-release (SAR) sequence mediates export and control of the phage P1 endolysin Proc Natl Acad Sci USA 2004, 101:6415-6420.

110. Park T, Struck DK, Dankenbring CA, Young R. The Pinholin of Lambdoid Phage 21: Control of Lysis by Membrane Depolarization. Journal of Bacteriology 2007, 189(24):9135-9139.

111. Pang T, Savva CG, Fleming KG, Struck DK, Young R. Structure of the lethal phage pinhole. Proceedings of the National Academy of Sciences 2009, 106(45):18966-18971.

112. Sun QK, G. F.; Arulandu, A.;Xu, M.; Young, R.; Sacchettini, J. C. Regulation of the phage 21 endolysin by topological dynamics and membrane proximity. In.; 2009.

113. Sun Q, Kuty GF, Arockiasamy A, Xu M, Young R, Sacchettini JC. Regulation of a muralytic enzyme by dynamic membrane topology. Nat Struct Mol Biol 2009, 16(11):1192-1194.

114. Berry J, Rajaure M, Pang T, Young R. The Spanin Complex Is Essential for Lambda Lysis. Journal of Bacteriology 2012, 194(20):5667-5674.

115. Berry J, Summer EJ, Struck DK, Young R. The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. Molecular Microbiology 2008, 70(2):341-351.

116. Summer EJ, Berry J, Tran TAT, Niu L, Struck DK, Young R. Rz/Rz1 Lysis Gene Equivalents in Phages of Gram-negative Hosts. Journal of Molecular Biology 2007, 373(5):1098-1112.

117. Rajaure M, Berry J, Kongari R, Cahill J, Young R. Membrane fusion during phage lysis. Proceedings of the National Academy of Sciences 2015, 112(17):5497-5502.

118. Grayhack EJ, Yang X, Lau LF, Roberts JW. Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. Cell 1985, 42(1):259-269.

119. Remaut E, Stanssens P, Fiers W. Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. Gene 1981, 15(1):81-93.

120. Bläsi U, Nam K, Hartz D, Gold L, Young R. Dual translational initiation sites control function of the lambda S gene. The EMBO Journal 1989, 8(11):3501-3510.

121. Graschopf A, Bläsi U. Molecular function of the dual-start motif in the λ S holin. Molecular Microbiology 1999, 33(3):569-582.

122. Gründling A, Bläsi U, Young R. Biochemical and Genetic Evidence for Three Transmembrane Domains in the Class I Holin, λ S. Journal of Biological Chemistry 2000, 275(2):769-776.

123. To KH, Young R. Probing the Structure of the S105 Hole. Journal of Bacteriology 2014, 196(21):3683-3689.

124. Bläsi U, Chang CY, Zagotta MT, Nam KB, Young R. The lethal lambda S gene encodes its own inhibitor. The EMBO Journal 1990, 9(4):981-989.

125. Gründling A, Smith DL, Bläsi U, Young R. Dimerization between the Holin and Holin Inhibitor of Phage λ. Journal of Bacteriology 2000, 182(21):6075-6081.

126. Oliveira H, Melo LDR, Santos SB, Nóbrega FL, Ferreira EC, Cerca N, Azeredo J, Kluskens LD. Molecular Aspects and Comparative Genomics of Bacteriophage Endolysins. Journal of Virology 2013, 87(8):4558-4570.

127. Young R, Wang I-N, Roof WD. Phages will out: strategies of host cell lysis. Trends in Microbiology 2000, 8(3):120-128.

128. Zhang N, Young R. Complementation and characterization of the nested *Rz* and *Rz1* reading frames in the genome of bacteriophage lambda. Molecular and General Genetics 1999, 262(4-5):659-667.

129. Berry JD, Rajaure M, Young R. Spanin function requires subunit homodimerization through intermolecular disulfide bonds. Molecular Microbiology 2013, 88(1):35-47.

130. Young R, Way J, Way S, Yin J, Syvanen M. Transposition mutagenesis of bacteriophage lambda: A new gene affecting cell lysis. Journal of Molecular Biology 1979, 132(3):307-322.

131. Hanych B, Kędzierska S, Walderich B, Uznański B, Taylor A. Expression of the Rz gene and the overlapping Rz1 reading frame present at the right end of the bacteriophage lambda genome. Gene 1993, 129(1):1-8.

132. Kędzierska S, Wawrzynów A, Taylor A. The Rz1 gene product of bacteriophage lambda is a lipoprotein localized in the outer membrane of Escherichia coli. Gene 1996, 168(1):1-8.

133. Terada M, Kuroda T, Matsuyama S-i, Tokuda H. Lipoprotein Sorting Signals Evaluated as the LolA-dependent Release of Lipoproteins from the Cytoplasmic Membrane of Escherichia coli. Journal of Biological Chemistry 2001, 276(50):47690-47694.

134. Cahill J, Rajaure M, O'Leary C, Sloan J, Marrufo A, Holt A, Kulkarni A, Hernandez O, Young R. Genetic Analysis of the Lambda Spanins Rz and Rz1: Identification of Functional Domains. G3 2017, 7(2):741-753.

135. Cahill J, Rajaure M, Holt A, Moreland R, O'Leary C, Kulkarni A, Sloan J, Young R. Suppressor analysis of the fusogenic lambda spanins. J Virol 2017.

136. Osborn MJ, Munson R. Separation of the inner (cytoplasmic) and outer membranes of gram-negative bacteria. In: *Methods in Enzymology.* vol. 31: Academic Press; 1974: 642-653.

137. Jahn R, Lang T, Südhof TC. Membrane Fusion. Cell 2003, 112(4):519-533.

138. Martens S, McMahon HT. Mechanisms of membrane fusion: disparate players and common principles. Nature Reviews Molecular Cell Biology 2008, 9:543.

139. Bour BA, Chakravarti M, West JM, Abmayr SM. Drosophila SNS, a member of the immunoglobulin superfamily that is essential for myoblast fusion. Genes & Development 2000, 14(12):1498-1511.

140. Merlini L, Dudin O, Martin SG. Mate and fuse: how yeast cells do it. Open Biology 2013, 3(3):130008.

141. Bonifacino JS, Glick BS. The Mechanisms of Vesicle Budding and Fusion. Cell 2004, 116(2):153-166.

142. Harrison SC. Viral membrane fusion. Nature structural & molecular biology 2008, 15(7):690-698.

143. Rilfors L, Lindblom G, Wieslander Å, Christiansson A. Lipid Bilayer Stability in Biological Membranes. In: *Membrane Fluidity.* Edited by Kates M, Manson LA. Boston, MA: Springer US; 1984: 205-245.

144. Chernomordik LV, Kozlov MM. Protein-Lipid Interplay in Fusion and Fission of Biological Membranes. Annual Review of Biochemistry 2003, 72(1):175-207.

145. Kozlovsky Y, Kozlov MM. Stalk model of membrane fusion: solution of energy crisis. Biophysical Journal 2002, 82(2):882-895.

146. Markin VS, Albanesi JP. Membrane fusion: stalk model revisited. Biophysical Journal 2002, 82(2):693-712.

147. Kuzmin PI, Zimmerberg J, Chizmadzhev YA, Cohen FS. A quantitative model for membrane fusion based on low-energy intermediates. Proceedings of the National Academy of Sciences of the United States of America 2001, 98(13):7235-7240.

148. Kozlov MM, Chernomordik LV. A mechanism of protein-mediated fusion: coupling between refolding of the influenza hemagglutinin and lipid rearrangements. Biophysical Journal 1998, 75(3):1384-1396.

149. Martens S, Kozlov MM, McMahon HT. How Synaptotagmin Promotes Membrane Fusion. Science 2007, 316(5828):1205.

150. Jackson MB, Chapman ER. Fusion pores and fusion machines in Ca2+ triggered exocytosis. Annual Review of Biophysics and Biomolecular Structure 2006, 35(1):135-160.

151. Earp LJ, Delos SE, Park HE, White JM. The Many Mechanisms of Viral Membrane Fusion Proteins. In: *Membrane Trafficking in Viral Replication.* Edited by Marsh M. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005: 25-66.

152. White JM, Delos SE, Brecher M, Schornberg K. Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. Critical reviews in biochemistry and molecular biology 2008, 43(3):189-219.

153. Kielian M, Rey FA. Virus membrane-fusion proteins: more than one way to make a hairpin. Nat Rev Micro 2006, 4(1):67-76.

154. Kielian M. Mechanisms of Virus Membrane Fusion Proteins. Annual Review of Virology 2014, 1(1):171-189.

155. Kielian M, Klimjack MR, Ghosh S, Duffus WA. Mechanisms of mutations inhibiting fusion and infection by Semliki Forest virus. The Journal of Cell Biology 1996, 134(4):863.

156. Skehel JJ, Wiley DC. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. Annual Review of Biochemistry 2000, 69(1):531-569.

157. Stiasny K, Kössl C, Lepault J, Rey FA, Heinz FX. Characterization of a Structural Intermediate of Flavivirus Membrane Fusion. PLoS Pathogens 2007, 3(2):e20.

158. Kozlov MM, McMahon HT, Chernomordik LV. Protein-driven membrane stresses in fusion and fission. Trends in biochemical sciences 2010, 35(12):699-706.

159. Cohen FS, Melikyan GB. The Energetics of Membrane Fusion from Binding, through Hemifusion, Pore Formation, and Pore Enlargement. The Journal of Membrane Biology 2004, 199(1):1-14.

160. Wiley DC, Skehel JJ. The Structure and Function of the Hemagglutinin Membrane Glycoprotein of Influenza Virus. Annual Review of Biochemistry 1987, 56(1):365-394.

161. Thoennes S, Li Z-N, Lee B-J, Langley WA, Skehel JJ, Russell RJ, Steinhauer DA. Analysis of Residues Near the Fusion Peptide in the Influenza Hemagglutinin Structure for Roles in Triggering Membrane Fusion. Virology 2008, 370(2):403-414.

162. Leikina E, Mittal A, Cho M-S, Melikov K, Kozlov MM, Chernomordik LV. Influenza Hemagglutinins Outside of the Contact Zone Are Necessary for Fusion Pore Expansion. Journal of Biological Chemistry 2004, 279(25):26526-26532.

163. Li L, Jose J, Xiang Y, Kuhn RJ, Rossmann MG. Structural Changes of Envelope Proteins During Alphavirus Fusion. Nature 2010, 468(7324):705-708.

164. Roussel A, Lescar J, Vaney M-C, Wengler G, Wengler G, Rey FA. Structure and Interactions at the Viral Surface of the Envelope Protein E1 of Semliki Forest Virus. Structure 2006, 14(1):75-86.

165. Gibbons DL, Vaney M-C, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA. Conformational change and protein–protein interactions of the fusion protein of Semliki Forest virus. Nature 2004, 427:320.

166. Sjöberg M, Lindqvist B, Garoff H. Activation of the Alphavirus Spike Protein Is Suppressed by Bound E3. Journal of Virology 2011, 85(11):5644-5650.

167. Antti Salminen JMW, Mario Lobigs, Peter Liljestrom, and Henrik Garoff. Membrane fusion process of Semliki Forest virus. II: Cleavage-dependent reorganization of the spike protein complex controls virus entry. The Journal of Cell Biology 1992, 116(2):349-357.

168. Top D, Read JA, Dawe SJ, Syvitski RT, Duncan R. Cell-Cell Membrane Fusion Induced by p15 Fusion-associated Small Transmembrane (FAST) Protein Requires a Novel Fusion Peptide Motif Containing a Myristoylated Polyproline Type II Helix. The Journal of Biological Chemistry 2012, 287(5):3403-3414.

169. Payne K, Sun Q, Sacchettini J, Hatfull GF. Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. Mol Micro 2009, 73(3):367-381.

170. Gil F, Grzegorzewicz AE, Catalão MJ, Vital J, McNeil MR, Pimentel M. Mycobacteriophage Ms6 LysB specifically targets the outer membrane of Mycobacterium smegmatis. Microbiology 2010, 156(Pt 5):1497-1504.

171. Yao Z, Kahne D, Kishony R. Distinct Single-Cell Morphological Dynamics under Beta-Lactam Antibiotics. Molecular Cell 2012, 48(5):705-712.

172. Rajaure M, Berry J, Kongari R, Cahill J, Young R. Membrane fusion during phage lysis. Proc Natl Acad Sci U S A 2015, 112(17):5497-5502.

173. Hatfull GF. Bacteriophage Genomics. Curr Opin Microbiol 2008, 11(5):447-453.

174. Perez Sepulveda B, Redgwell T, Rihtman B, Pitt F, Scanlan DJ, Millard A. Marine phage genomics: the tip of the iceberg. FEMS Microbiol Lett 2016, 363(15):fnw158.

175. Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. Nucleic Acids Res 1998, 26(2):544-548.

176. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 1999, 27(23):4636-4641.

177. Besemer J, Borodovsky M. Heuristic approach to deriving models for gene finding. Nucleic Acids Res 1999, 27(19):3911-3920.

178. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 2001, 29(12):2607-2618.

179. Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 1998, 26(4):1107-1115.

180. Chirico N, Vianelli A, Belshaw R. Why genes overlap in viruses. Proc R Soc Lond [Biol] 2010, 277(1701):3809-3817.

181. Mijalis E, Rasche E. Find Spanin Workflow https://cpt.tamu.edu/galaxy-pub/u/elenimijalis/w/imported-findspanin-v20-apollo Accessed 1st Jan 2018

182. Spanin Database. https://cpt.tamu.edu/spanindb/#/phages Accessed.1st Jan 2018

183. Graham LL, Harris R, Villiger W, Beveridge TJ. Freeze-substitution of gram-negative eubacteria: general cell morphology and envelope profiles. J Bacteriol 1991, 173(5):1623-1633.

184. BLASTCLUST tool. https://cpt.tamu.edu/galaxy/root?tool_id=blastclust Accessed 1st Jan 2018

185. NCBI. Documentation of the BLASTCLUST-algorithm. ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html Accessed 1st Jan 2018

186. Pearson WR. Selecting the Right Similarity-Scoring Matrix. Curr Protoc Bioinformatics 2013, 43:3.5.1-3.5.9.

187. Schaefer MH, Wanker EE, Andrade-Navarro MA. Evolution and function of CAG/polyglutamine repeats in protein–protein interaction networks. Nucleic Acids Res 2012, 40(10):4273-4287.

188. Ahern SJ, Das M, Bhowmick TS, Young R, Gonzalez CF. Characterization of Novel Virulent Broad-Host-Range Phages of Xylella fastidiosa and Xanthomonas. Journal of Bacteriology 2014, 196(2):459-471.

189. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping Genes Produce Proteins with Unusual Sequence Properties and Offer Insight into De Novo Protein Creation. J Virol 2009, 83(20):10719-10736.

190. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK. TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder. Protein and peptide letters 2008, 15(9):956-963.

191. von Heijne G. The structure of signal peptides from bacterial lipoproteins. Protein Engineering 1989, 2(7):531-534.

192. Babu MM, Priya ML, Selvan AT, Madera M, Gough J, Aravind L, Sankaran K. A Database of Bacterial Lipoproteins (DOLOP) with Functional Assignments to Predicted Lipoproteins. Journal of Bacteriology 2006, 188(8):2761-2773.

193. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator. Genome Research 2004, 14(6):1188-1190.

194. Firth AE, Brown CM. Detecting overlapping coding sequences with pairwise alignments. Bioinformatics 2005, 21(3):282-292.

195. Firth AE, Brown CM. Detecting overlapping coding sequences in virus genomes. BMC Bioinformatics 2006, 7:75-75.

196. LipoSearch Tool. https://cpt.tamu.edu/galaxy-pub/u/ramchander-r-kongari/w/liposearch-gff-v2 Accessed 1st Jan 2018

197. Hernandez-Morales AC, Lessor LL, Wood TL, Migl D, Mijalis EM, Russell WK, Young RF, Gill JJ. Genomic and Biochemical Characterization of Acinetobacter Podophage Petty Reveals a Novel Lysis Mechanism and Tail-Associated Depolymerase Activity. J Virol 2018.

198. Lai M-J, Lin N-T, Hu A, Soo P-C, Chen L-K, Chen L-H, Chang K-C. Antibacterial activity of Acinetobacter baumannii phage ϕAB2 endolysin (LysAB2) against both Gram-positive and Gram-negative bacteria. Appl Microbiol Biotechnol 2011, 90(2):529-539.

199. Lood R, Winer BY, Pelzek AJ, Diez-Martinez R, Thandar M, Euler CW, Schuch R, Fischetti VA. Novel Phage Lysin Capable of Killing the Multidrug-Resistant Gram-Negative Bacterium Acinetobacter baumannii in a Mouse Bacteremia Model. Antimicrob Agents Chemother 2015, 59(4):1983-1991.

200. Jung CH, Yang Y-S, Kim J-S, Shin J-I, Jin Y-S, Shin JY, Lee JH, Chung KM, Hwang JS, Oh JM *et al*. A search for synthetic peptides that inhibit soluble N-ethylmaleimide sensitive-factor attachment receptor-mediated membrane fusion. FEBS J 2008, 275(12):3051-3063.

201. Sia SK, Carr PA, Cochran AG, Malashkevich VN, Kim PS. Short constrained peptides that inhibit HIV-1 entry. Proc Natl Acad Sci U S A 2002, 99(23):14664-14669.

202. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res 2015, 43(W1):W389-W394.

203. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science 1991, 252(5009):1162.

204. Seydel A, Gounon P, Pugsley AP. Testing the '+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. Molecular Microbiology 1999, 34(4):810-821.

205. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 2011, 8(10):785-786.

206. Juncker AS, Willenbrock H, von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Science : A Publication of the Protein Society 2003, 12(8):1652-1662.

207. Hussain M, Ichihara S, Mizushima S. Accumulation of glyceride-containing precursor of the outer membrane lipoprotein in the cytoplasmic membrane of Escherichia coli treated with globomycin. Journal of Biological Chemistry 1980, 255(8):3707-3712.

208. Inukai M, Takeuchi M, Shimizu K. Effects of Globomycin on the Morphology of Bacteria and the Isolation of Resistant Mutants. Agricultural and Biological Chemistry 1984, 48(2):513-518.

209. Strych U, Penland RL, Jimenez M, Krause KL, Benedik MJ. Characterization of the alanine racemases from two Mycobacteria. FEMS Microbiology Letters 2001, 196(2):93-98.

210. Andersson H, von Heijne G. Sec dependent and sec independent assembly of E. coli inner membrane proteins: the topological rules depend on chain length. The EMBO Journal 1993, 12(2):683-691.

211. Andersson H, von Heijne G. Membrane protein topology: effects of delta mu H+ on the translocation of charged residues explain the 'positive inside' rule. The EMBO Journal 1994, 13(10):2267-2272.

212. Aschtgen M-S, Zoued A, Lloubès R, Journet L, Cascales E. The C-tail anchored TssL subunit, an essential protein of the enteroaggregative Escherichia coli Sci-1 Type VI secretion system, is inserted by YidC. MicrobiologyOpen 2012, 1(1):71-82.

213. Borgese N, Righi M. Remote Origins of Tail-Anchored Proteins. Traffic 2010, 11(7):877-885.

214. Hinderhofer M, Walker CA, Friemel A, Stuermer CAO, Möller HM, Reuter A. Evolution of prokaryotic SPFH proteins. BMC Evolutionary Biology 2009, 9(1):10.

215. Walker CA, Hinderhofer M, Witte DJ, Boos W, Möller HM. Solution structure of the soluble domain of the NfeD protein YuaF from Bacillus subtilis. Journal of Biomolecular NMR 2008, 42(1):69-76.

216. Theobald DL, Mitton-Fry RM, Wuttke DS. Nucleic acid recognition by OB-fold proteins. Annual review of biophysics and biomolecular structure 2003, 32:115-133.

217. Haeusser DP, Hoashi M, Weaver A, Brown N, Pan J, Sawitzke JA, Thomason LC, Court DL, Margolin W. The Kil Peptide of Bacteriophage λ Blocks Escherichia coli Cytokinesis via ZipA-Dependent Inhibition of FtsZ Assembly. PLOS Genetics 2014, 10(3):e1004217.

218. To KH, Dewey J, Weaver J, Park T, Young R. Functional Analysis of a Class I Holin, P2 Y. Journal of Bacteriology 2013, 195(6):1346-1355.

219. Bonovich MT, Young R. Dual start motif in two lambdoid S genes unrelated to lambda S. J Bacteriol 1991, 173(9):2897-2905.

220. Powell BS, Rivas MP, Court DL, Nakamura Y, Turnbough CL. Rapid confirmation of single copy lambda prophage integration by PCR. Nucleic Acids Research 1994, 22(25):5765-5766.

221. Park T, Struck DK, Deaton JF, Young R. Topological dynamics of holins in programmed bacterial lysis. Proceedings of the National Academy of Sciences of the United States of America 2006, 103(52):19713-19718.

222. Giugliarelli A, Urbanelli L, Ricci M, Paolantoni M, Emiliani C, Saccardi R, Mazzanti B, Lombardini L, Morresi A, Sassi P. Evidence of DMSO-Induced Protein Aggregation in Cells. The Journal of Physical Chemistry A 2016, 120(27):5065-5070.

223. Rietsch A, Beckwith J. The genetics of disulfide bond metabolism. Annual Review of Genetics 1998, 32(1):163-184.

224. KuppusamyKrishnan H, Chau LM, Moreno-Hagelsieb G, Saier MH. Analysis of 58 families of Holins Using A Novel Program, PhyST: PhyST evaluates familial protein characteristics and their organismal sources. Journal of molecular microbiology and biotechnology 2016, 26(6):381-388.

225. Vidová B, Šramková Z, Tišáková L, Oravkinová M, Godány A. Bioinformatics analysis of bacteriophage and prophage endolysin domains. Biologia 2014, 69(5):541-556.

226. Xu M, Arulandu A, Struck DK, Swanson S, Sacchettini JC, Young R. Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science 2005, 307:113-117.

227. Xu M, Struck DK, Deaton J, Wang I-N, Young R. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proceedings of the National Academy of Sciences of the United States of America 2004, 101(17):6415-6420.

228. Chanel-Vos C, Kielian M. Second-Site Revertants of a Semliki Forest Virus Fusion-Block Mutation Reveal the Dynamics of a Class II Membrane Fusion Protein. Journal of Virology 2006, 80(12):6115-6122.

229. Spudich GM, Fernandez D, Zhou XR, Christie PJ. Intermolecular disulfide bonds stabilize VirB7 homodimers and VirB7/VirB9 heterodimers during biogenesis of the Agrobacterium tumefaciens T-complex transport apparatus. Proceedings of the National Academy of Sciences of the United States of America 1996, 93(15):7512-7517.

230. Nagano K, Read EK, Murakami Y, Masuda T, Noguchi T, Yoshimura F. Trimeric Structure of Major Outer Membrane Proteins Homologous to OmpA in Porphyromonas gingivalis. Journal of Bacteriology 2005, 187(3):902-911.

231. Klimke WA, Rypien CD, Klinger B, Kennedy RA, Rodriguez-Maillard JM, Frost LS. The mating pair stabilization protein, TraN, of the F plasmid is an outer-membrane protein with two regions that are important for its function in conjugation. Microbiology 2005, 151(11):3527-3540.

232. Usón I, Patzer SI, Rodríguez DD, Braun V, Zeth K. The crystal structure of the dimeric colicin M immunity protein displays a 3D domain swap. Journal of Structural Biology 2012, 178(1):45-53.

233. Diao J, Ishitsuka Y, Bae W-R. Single-molecule FRET study of SNARE-mediated membrane fusion. Bioscience reports 2011, 31(6):457-463.

234. Brooks A, Shoup D, Kustigian L, Puchalla J, Carr CM, Rye HS. Single Particle Fluorescence Burst Analysis of Epsin Induced Membrane Fission. PLoS ONE 2015, 10(3):e0119563.

235. Keidel A, Bartsch TF, Florin E-L. Direct observation of intermediate states in model membrane fusion. Scientific Reports 2016, 6:23691.

236. Zhang X, Ma L, Zhang Y. High-Resolution Optical Tweezers for Single-Molecule Manipulation. The Yale Journal of Biology and Medicine 2013, 86(3):367-383.